

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE CIÊNCIAS EXATAS E DA TERRA
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA



Programa de Pós-Graduação
em Química



Novas metodologias espectroscópicas e eletroquímica aplicando
classificação multivariada em sistemas biológicos

Leomir Aires Silva de Lima

Tese de Doutorado
Natal/RN, março de 2018

Leomir Aires Silva de Lima

**NOVAS METODOLOGIAS ESPECTROSCÓPICAS E
ELETROQUÍMICA APLICANDO CLASSIFICAÇÃO MULTIVARIADA
EM SISTEMAS BIOLÓGICOS.**

Tese de doutorado submetida ao Programa de Pós-Graduação em Química – PPGQ – do Instituto de Química da Universidade Federal do Rio Grande do Norte em cumprimento as exigências para a obtenção do título Doutor em Química.

Orientador: Prof.^o Dr. Kássio Michell Gomes de Lima
Co-orientador: Prof. Dr. Luiz da Silva Gasparotto

**NATAL-RN
2018**

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Lima, Leomir Aires Silva de.

Novas metodologias espectroscópicas e eletroquímica aplicando
classificação multivariada em sistemas biológicos / Leomir Aires
Silva de Lima. - 2018.

124 f.: il.

Tese (doutorado) - Universidade Federal do Rio Grande do
Norte, Centro de Ciências Exatas e da Terra, Programa de Pós-
Graduação em Química. Natal, RN, 2018.

Orientador: Prof. Dr. Kássio Michell Gomes de Lima.

Coorientador: Prof. Dr. Luiz Henrique da Silva Gasparotto.

1. Fluorescência - Tese. 2. Voltametria - Tese. 3. FTIR -
Tese. 4. Sarcopenia - Tese. 5. Entomotoxicologia - Tese. 6.
Análise Multivariada - Tese. I. Lima, Kássio Michell Gomes de.
II. Gasparotto, Luiz Henrique da Silva. III. Título.

RN/UF/BCZM

CDU 543.5

Leomir Aires Silva de Lima

NOVAS METODOLOGIAS ESPECTROSCÓPICAS E ELETROQUÍMICA
APLICANDO CLASSIFICAÇÃO MULTIVARIADA EM SISTEMAS BIOLÓGICOS

Tese apresentada ao Programa de Pós-graduação em Química da Universidade Federal do Rio Grande do Norte, em cumprimento às exigências para obtenção do título de Doutor em Química.

Aprovada em: 16 de março de 2018

Comissão Examinadora:



Dr. Kássio Michell Gomes de Lima – UFRN (orientador)



Dr. Fabrício Gava Menezes – UFRN



Dra. Renata Antonaci Gama – UFRN



Dr. Edvan Cirino da Silva – UFPB



Dr. Sherlan Guimarães Lemos – UFPB

AGRADECIMENTOS

À Deus acima de tudo.

Aos meus pais Baltazar Francisco de Lima e Terezinha de Jesus Silva de Lima, por todo apoio e por me permitir o acesso à educação e fazer disso sua prioridade.

A minha esposa Euclécia Elaine da Silva por toda compreensão, apoio e dedicação durante essa longa caminhada, ela mais do que qualquer outro sabe o quanto foi árduo chegar até aqui.

Aos meus avós paternos Francisca Canela da Silva (*In memoriam*) e Manoel Francisco de Lima (*In memoriam*) e avós maternos Luiza Bezerra da Silva (*In memoriam*) e Manoel Bezerra da Silva (*In memoriam*) pelo pouco tempo que tive a felicidade de conviver e hoje vos guardam com muito carinho na memória.

Aos meus sogros, Cleide e Eugênio, meus irmãos, sobrinhos, tios, tias, primos, primas, cunhado, cunhadas, compadres e afilhadas pelo carinho e companheirismo.

Ao meu orientador, Prof^o. Dr. Kássio Michell Gomes de Lima, por acreditar em meu potencial, pelos ensinamentos e confiança em mim depositada, como também por ter me dado à chance de conhecer essa ciência linda chamada quimiometria.

Aos meus sócios Camilo, Fernanda e Heloisa pela parceria cotidiana, pela força recíproca e pela oportunidade de juntos podermos materializar um sonho chamado ChemoVector.

A todo o grupo de química biológica e quimiometria (QbQ) pelas horas de dedicação e troca de conhecimento.

Aos membros do laboratório de insetos e vetores (LIVe), em especial a coordenadora prof. Dra. Renata Gama pela parceria e trocas de conhecimento e a mestrando Jéssica Jales pela parceria em diversos trabalhos.

Ao prof. Dr. Sherlan G. Lemos por ter aberto as portas de seu laboratório e permitido usar a instrumentação eletroquímica essencial para a conclusão dessa tese e contribuição no desenvolvimento desse trabalho.

A CAPES pela bolsa concedida e financiamento do trabalho.

RESUMO

Esta tese demonstra o desenvolvimento de novas metodologias instrumentais como: espectroscopias no infravermelho próximo e médio, de fluorescência molecular e voltametria de pulso diferencial, combinadas a técnicas de classificação multivariada e seleção de variáveis, aplicadas a sistema biológico como tecido, larvas de mosca e plasma sanguíneo. Em um primeiro estudo, a espectroscopia NIR foi utilizada para obtenção de espectros de crânio de rato saudáveis e com defeito crítico. Modelos multivariados de classificação foram construídos, visando uma metodologia alternativa para estimar a taxa de sucesso do reparo ósseo através dos tratamentos com laser de baixa intensidade e colágeno. Os algoritmos PCA-LDA, SPA-LDA e GA-LDA foram aplicados como ferramentas de classificação e seus desempenhos comparados. De maneira geral, os modelos foram capazes de classificar o grupo controle dos demais mostrando eficiência na avaliação da cicatrização de feridas em defeito craniano de rato. Em um segundo estudo, a espectroscopia ATR-FTIR foi utilizada para obtenção de espectros de plasmas sanguíneos de idosos saudáveis e portadores de sarcopenia. Modelos multivariados de classificação foram construídos, visando uma metodologia de diagnóstico para triagem de idosos sarcopênicos. Os algoritmos PCA-LDA, SPA-LDA e GA-LDA foram aplicados como ferramentas de classificação e seus desempenhos comparados. De maneira geral, os resultados obtidos através do algoritmo de seleção de variáveis SPA-LDA e GA-LDA foram os mais satisfatórios alcançando sensibilidade de 86,7%, pois puderam ser relacionadas a grupos funcionais pertencentes a diferentes biomoléculas e possíveis biomarcadores. Em um terceiro estudo, a voltametria de pulso diferencial e matrizes de excitação/emissão de fluorescência molecular foram obtidas para extratos de larvas de mosca controle e contendo o fármaco flunitrazepam. Foi avaliado o desempenho dos algoritmos de classificação com métodos de seleção de variáveis e funções discriminante linear (sensibilidade de 90%) e quadrática (sensibilidade de 100%) para os voltamogramas e o algoritmo 2D-LDA (sensibilidade de 100%) comparado ao PARAFAC-LDA (sensibilidade de 40%) para classificação entre os grupos investigado. Os resultados sugerem que métodos espectroscópicos e eletroquímicos somados à análise multivariada são capazes de minerar informações relevantes a cerca do entendimento da dinâmica biológica de diversos organismos, como também contribuir com o sistema de saúde pública na identificação de sarcopenia em idosos e classificar insetos contendo drogas.

Palavras-chave: Fluorescência. FT-IR. NIR. Sarcopenia. Voltametria. Entomotoxicologia. Análise Multivariada.

ABSTRACT

This thesis demonstrates the development of new instrumental methodologies such as: near and medium infrared spectroscopy, molecular fluorescence and differential pulse voltammetry, combined with multivariate classification and variable selection techniques applied to biological system such as tissue, fly larvae and blood plasma. In a first study, NIR spectroscopy was used to obtain healthy and critically defective rat skull spectra. Multivariate classification models were constructed, aiming at an alternative methodology to estimate the success rate of bone repair through low intensity laser and collagen treatments. The algorithms PCA-LDA, SPA-LDA and GA-LDA were applied as classification tools and their comparative performances. In general, the models were able to classify the control group of the others showing efficiency in the evaluation of wound healing in rat cranial defect. In a second study, ATR-FTIR spectroscopy was used to obtain blood plasma spectra of healthy elderly patients with sarcopenia. Multivariate classification models were constructed, aiming at a diagnostic methodology for screening of sarcopenic elderly. The algorithms PCA-LDA, SPA-LDA and GA-LDA were applied as classification tools and their comparative performances. In general, the results obtained through the SPA-LDA and GA-LDA variable selection algorithm were the most satisfactory, achieving sensitivity of 86.7%, since they could be related to functional groups belonging to different biomolecules and possible biomarkers. In a third study, differential pulse voltammetry and molecular fluorescence excitation / emission matrices were obtained for extracts of control fly larvae and containing the drug flunitrazepam. The performance of classification algorithms was evaluated using linear discriminant (90% sensitivity) and quadratic (100% sensitivity) methods for the voltamograms and the 2D-LDA algorithm (sensitivity of 100%) compared to PARAFAC - LDA (sensitivity of 40%) for classification between the groups investigated. The results suggest that spectroscopic and electrochemical methods added to the multivariate analysis are capable of mining relevant information about the understanding of the biological dynamics of several organisms, as well as contributing to the public health system in the identification of sarcopenia in the elderly and classify insects with drugs.

Keywords: Fluorescence. ATR-FTIR. NIR. Sarcopenia. Voltammetry. Entomotoxicology. Multivariate analysis.

LISTA DE ABREVIATURAS E SIGLAS

- ATR** – reflectância total atenuada (do inglês, *attenuated total reflection*)
- GC** – cromatografia gasosa (do inglês, *gas-chromatography*)
- DA** – análise discriminante (do inglês, *discriminant analysis*)
- DPV** – voltametria de pulso diferencial (do inglês, *differential pulse voltammetry*)
- EEM** – matriz de excitação/emissão (do inglês, *excitation/emission matrix*)
- FT** – transformada de Fourier (do inglês, *Fourier transform*)
- GA** – algoritmo genético (do inglês, *genetic algorithm*)
- IR** – infravermelho (do inglês, *infrared*)
- LC** – cromatografia líquida (do inglês, *liquid-chromatography*)
- LDA** – análise discriminante linear (do inglês, *linear discriminant analysis*)
- MALDI** – ionização e dessorção a laser assistida por matriz (do inglês, *Matrix Assisted Laser Desorption/Ionization*)
- MIR** – infravermelho médio (do inglês, *mid-infrared*)
- MS** – Espectrometria de massas (do inglês, *mass spectrometry*)
- NIR** – infravermelho próximo (do inglês, *near-infrared*)
- PARAFAC** – análise de fatores paralelos (do inglês, *parallel factor analysis*)
- QDA** – análise discriminante quadrática (do inglês, *quadratic discriminant analysis*)
- PC** – componente principal (do inglês, *principal component*)

SUMÁRIO

Capítulo 1	Introdução geral.....	10
Capítulo 2	Evaluation of the bony repair in rat cranial defect using near infrared reflectance spectroscopy and discriminant analysis. Leomir A. S. de Lima , Lana S. de S. Oliveira, Aurigena A. Araújo, Raimundo F. de A. Junior, Kássio M. G. Lima. <i>Biotechnol. Prog.</i> , 2017, Vol. 33, No. 4, 1160 – 1167.....	40
Capítulo 3	Ferramenta de rastreamento para sarcopenia em idosos: uma abordagem bioespectroscópica. Leomir A. S. de Lima , Daniele S. Pereira, Juliana F. de S. Barbosa, Aline N. F. Freire, Ricardo O. Guerra, Kássio M. G. de Lima. Manuscrito em fase de escrita.....	50
Capítulo 4	Identification using classification analysis of flunitrazepam in necrophagous larvae via differential pulse voltammetry and fluorescence excitation-emission matrix (EEM) spectroscopy. Leomir A. S. de Lima , Camilo L. M. Moraes, Jéssica T. Jales, Renata A. Gama, Sherlan G. Lemos, Kássio M. G. Lima. Manuscrito submetido à Journal of the Brazilian Chemical Society.....	73
Capítulo 5	Conclusão e perspectivas.....	101
Apêndice A	Near infrared spectroscopy as an emerging tool for forensic entomotoxicology. Leomir A. S. de Lima , Tainá C. Baia, Renata A. Gama, Luiz H. da S. Gasparotto, Kássio M.G. Lima. <i>NIRnews</i> , 2014, vol. 25, 8, 5 – 7.	103
Apêndice B	FTIR microspectroscopy coupled with variable selection methods for the identification of flunitrazepam in necrophagous flies. Tainá C. Baia, Renata A. Gama, Leomir A. S. de Lima , Kássio M. G. Lima. <i>Analytical Methods</i> , 2016, 8, 968–972	107
Apêndice C	MCR-ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel-diesel blends. Anne B. F. Câmara, Luciene S. de Carvalho, Camilo L. M. de Moraes, Leomir A. S. de Lima , Heloise O. M. de Araújo, Fernanda M. de Oliveira,	

Kássio M.G. de Lima. <i>Fuel</i> 210 (2017) 497–506.....	114
--	-----

Capítulo 1 – Introdução Geral

1. Organização da tese	09
2. Introdução	10
3. Objetivos	29
4. Metodologia.....	30
Referências	35

1 Organização da tese

A presente tese é apresentada em capítulos referentes aos trabalhos como primeiro autor, que relatam estudos focados no tema central desenvolvido com colaboradores do centro de biociências da UFRN. Ainda, nos apêndices são apresentados trabalhos realizados através de diferentes colaborações que ocorreram durante o período de doutoramento.

Capítulo 2 - *“Evaluation of the bony repair in rat cranial defect using near infrared reflectance spectroscopy and discriminant analysis”* (publicado no periódico Biotechnology Progress, DOI: 10.1002/btpr.2476) - relata uma aplicação de infravermelho próximo e algoritmos de classificação multivariada como PCA-LDA, SPA-LDA e GA-LDA para avaliar a taxa de recuperação óssea em crânio de rato e também como uma metodologia auxiliar ao método padrão que é o raio-X

Capítulo 3 - *“Ferramenta de triagem para sarcopenia em idosos: uma abordagem bioespectroscópica.”* (manuscrito em fase de escrita) - relata uma aplicação de espectroscopia no infravermelho e análise discriminante multivariada para rastreamento de sarcopenia em idosos utilizando o plasma sanguíneo. Trata-se de uma metodologia inovadora para a área da fisioterapia.

Capítulo 4 - *“Identification using classification analysis of flunitrazepam in necrophagous larvae via differential pulse voltammetry and fluorescence excitation-emission matrix (EEM) spectroscopy”* (manuscrito submetido à revista *Journal of the Brazilian Chemical Society*) - relata uma aplicação de voltametria de pulso diferencial e fluorescência molecular em conjunto com análise multivariada para classificação de moscas necrófagas contendo o fármaco flunitrazepam, com enfoque na área de entomotoxicologia forense.

Capítulo 5 - *“Conclusões e Perspectivas”* - apresenta de forma resumida os principais resultados alcançados dentre os estudos, sua concordância com os objetivos

propostos e a relevância para o cada subtema abordado, além de perspectivas para trabalhos futuros.

Apêndice A - “*Near infrared spectroscopy as an emerging tool for forensic entomotoxicology*” (publicado no periódico NIRnews, DOI: 10.1255/nirn.1489) - relata uma nova aplicação do uso do infravermelho próximo com métodos de seleção de variáveis para classificar contaminação de moscas por drogas.

Apêndice B - “FTIR microspectroscopy coupled with variable selection methods for the identification of flunitrazepam in necrophagous flies” (publicado no periódico Analytical Methods DOI: 10.1039/c5ay02342d) – Aplicação da espectroscopia ATR-FTIR em larvas de mosca para identificação de flunitrazepam, como uma nova metodologia para a área de perícia.

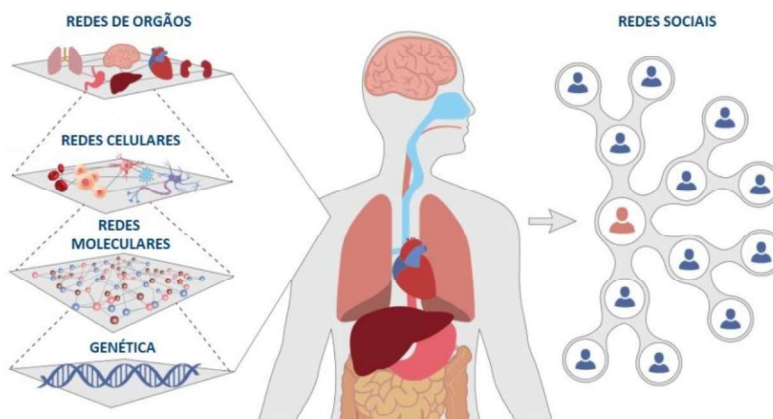
Apêndice C - “*MCR-ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel-diesel blends*” (publicado no periódico Fuel, DOI: 10.1016/j.fuel.2017.08.072) - retrata a aplicação de espectroscopia NIR/MIR para a quantificação e identificação de adulterantes em misturas de diesel-biodiesel utilizando de regressão como PLS e MCR-ALS.

2 Introdução

Historicamente, o entendimento sobre a biologia humana tem apresentado escassez de data e conteúdo informativo tornando rudimentar e incompleta a compreensão de processos biológicos complexos e análise de sistemas. A complexidade e variabilidade em nível celular, individual e populacional têm sido constantemente restringidas pela falta de tecnologias adequadas. Além disso, nossa compreensão da natureza dinâmica do metabolismo e da fisiologia humana como função do tempo, variando de minutos a anos, também é extremamente limitada. Isso se deve ao fato de haver substancial complexidade e variabilidade entre indivíduos e populações [1]. O ser humano apresenta cerca de 210 tipos celulares diferentes, 35 mil genes, 10 milhões de proteínas, entre 2 e 3 mil metabólitos[1,2]. Cada célula humana produz aproximadamente 95 trilhões de moléculas, 60 bilhões de proteínas, 2 trilhões de lipídeos, 60 bilhões de RNA [2]. Cada indivíduo produz informações bioquímicas particulares que podem ser influenciadas por fatores genéticos, ambientais e até culturais, tornando ainda mais árduo o entendimento da biologia destes animais [3] ilustrado pela Figura 1. Desta forma, informações acerca da

biologia humana requerem atenção, uma vez que remetem a individualidade e complexidade do ser humano.

Figura 1: Descrição da complexidade e fatores que influenciam a dinâmica molecular humana.



Fonte: Adaptado da referência [4]

Na tentativa de desvendar a complexidade e variabilidade humana, plataformas de conhecimento biológico foram e estão sendo desenvolvidas para análise de genoma [5] proteoma [6] e metaboloma [7] de matrizes biológicas complexas. Porém, tais plataformas geram conjuntos de dados maciços, analisados com tecnologias de ponta da bioinformática [3,5]. Além disso, as metodologias aplicadas para o sequenciamento e análise do genoma, identificação de todas as proteínas produzidas pelas células, de todos os metabólitos produzidos pelo corpo, bem como a correlação entre todas estas informações apresentam alto custo, requerem treinamento substancial para realização e, especialmente, análise dos dados produzidos [5,7,8].

As análises de sequenciamento de genoma empregaram tecnologias como o sequenciamento BAC-based, onde o genoma é fragmentado e posteriormente clonado no interior de bactérias para posterior análise em sequenciadores automatizados [5,8]. A plataforma proteômica conta com a realização de técnicas como Espectrometria de Massas, Captura Proteica por Anticorpos e análise de dados por ferramentas de bioinformática [9,10]. Já as análises metabolômicas são realizadas com a utilização de Espectrometria de ressonância magnética nuclear, espectroscopia do infravermelho com Transformada de Fourier (FTIR), Espectroscopia de massa acoplada com técnicas de separação, como cromatografia gasosa, cromatografia líquida em camada delgada [7].

Apesar da especificidade e sensibilidade elevadas das técnicas aplicadas na análise genômica, proteômica e metabolômica, o custo para realização e manutenção dos laboratórios, bem como do treinamento de indivíduos é substancial e acabam sendo limitadores quando aplicados a grande volumes de análise. Além disso, metodologias mais simples de análise e manutenção dos conjuntos de dados extraídos destas plataformas são necessárias para facilitar a compreensão das informações produzidas.

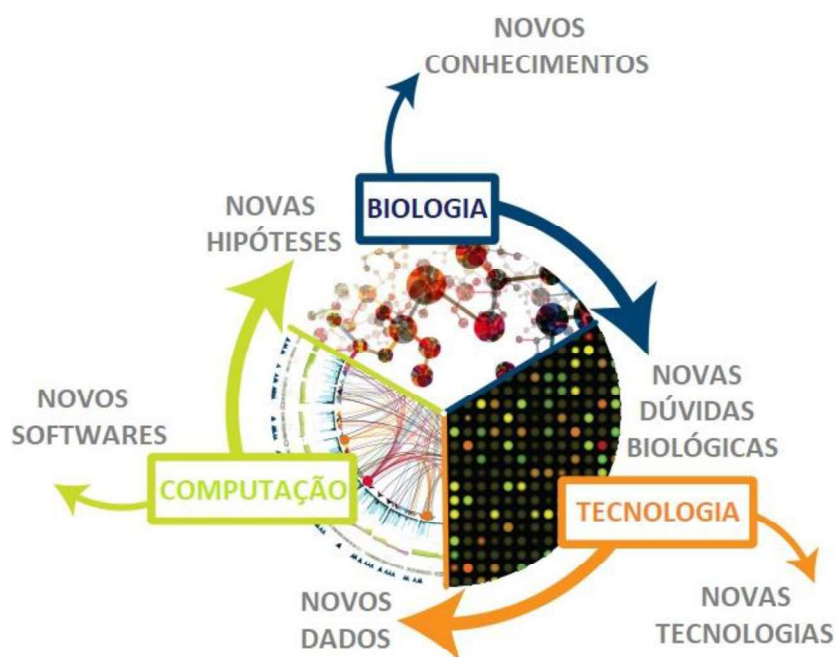
O avanço tecnológico tanto de hardwares quanto de softwares tem impulsionado fortemente às ciências, no qual a química tem sido amplamente beneficiada, principalmente no tocante à química analítica. A instrumentação analítica de alta precisão como LC-MS, CG-MS, MALDI TOF dentre outros tem sido importante aliada da biologia, ciências de materiais, medicina, ciências farmacêuticas dentro várias outras incapaz de serem totalmente mencionadas [6,7]. No entanto, essas instrumentações não estão facilmente à disposição em todas as realidades, principalmente em países subdesenvolvidos.

As espectroscopias no infravermelho e de fluorescência molecular, assim como técnicas eletroquímicas como a voltametria de pulso diferencial, de forma crescente vêm se inserindo em contextos biológicos, uma vez que permitem a obtenção de informação química sobre um grande e variado número de moléculas envolvidas direta ou indiretamente em processos biológicos [11,12]. As instrumentações espectroscópicas e eletroquímicas são potenciais ferramentas para serem utilizadas na identificação de drogas em humanos e insetos, avaliar a capacidade de recuperação de tecido biológico, na identificação de doenças, especialmente no que diz respeito a métodos de rastreamento. As vantagens dessas técnicas são rapidez e simplicidade na aquisição de dados, mínima ou nenhum processamento de amostra e baixo custo de materiais e equipamentos. Porém, tais técnicas geram dados de alta complexidade e baixa especificidade, especialmente quando se analisa amostras biológicas reais dadas à complexidade intrínseca presente neste tipo de matriz e o fato de que inúmeros interferentes (presença de moléculas irrelevantes à investigação) estão gerando sinais simultaneamente com os analito alvo. Desta forma, ferramentas computacionais multivariadas são de extrema importância para o sucesso dessas metodologias [13].

As ferramentas computacionais são capazes de filtrar as informações químicas pertinentes ao sistema em estudo, minimizando a influência de interferentes físicos e informações desnecessárias para o entendimento do sistema investigado. Com o propósito

de estudar sistemas biológicos, por exemplo, diversos algoritmos podem ser utilizados tanto para pré-processamento de dados quanto para classificação de amostras em grupos que representem ausência ou presença de biomarcadores e que possam estar associados a uma patologia [14], a presença de uma droga ilícita, ao aumento, diminuição ou alteração de biomoléculas importantes ao metabolismo animal ou vegetal. Alguns dos algoritmos mais frequentemente utilizados para dados de 1º ordem incluem o método de redução de dimensionalidade (PCA) e seleção de variáveis (SPA e GA) e para métodos de 2º ordem o PARAFAC e 2D associados a classificadores como LDA, QDA ou SVM. Na Figura 2 podemos observar a interdependência entre os sistemas biológicos, a instrumentação analítica e métodos quimiométricos que serve como força motriz para o desenvolvimento de novas pesquisas de caráter multidisciplinar capazes de dar um olhar diferente para áreas e problemas engessados no tempo.

Figura 2: Representação da interação entre biologia, tecnologia e computação aplicada a geração de novos conhecimentos.



Fonte: Adaptado da referência [4]

Desta forma, as técnicas espectroscópica e eletroquímica associadas à análise multivariada são ferramentas fortemente capazes de fornecer informações relevantes a cerca do entendimento e comportamento de sistemas biológicos complexos e capazes de atuar como alternativa de fácil implementação e baixo custo. Sendo assim, contribuindo

para uma aproximação das ciências químicas e biológicas e favorecendo o desenvolvimento e criação de novos métodos multidisciplinar.

2.1 Técnicas Instrumentais

2.1.1 Espectroscopia no infravermelho próximo (NIR)

A primeira faixa do espectro não visível a ser descoberta foi à região do NIR, porém sua utilização como fonte de informação analítica só ocorreu muito tempo depois sendo utilizada inicialmente em pesquisas na área de alimentos, no qual na década após os anos 1980 os números de publicações cresceram de 255 para 1000 [11]. A radiação infravermelha compreende a faixa de 780 a 100000 nm. O espectro do infravermelho é dividido em infravermelho próximo (*NearInfraRed*– NIR), infravermelho médio (*MiddleInfrared*– MIR) e infravermelho distante (*FAR Infrared*– FAR). A Tabela 1 abaixo apresenta os limites aproximados para cada região.

Tabela 1 – regiões espectrais do infravermelho

Região	Intervalos em números de ondas (cm⁻¹)	Intervalo em comprimento de onda, λ [nm]
Próximo (NIR)	12800 – 4000	780 – 2500
Médio (MID)	4000 – 200	2500 – 5000
Distante (FAR)	200 – 10	5000 – 10000

Fonte: Referência [15]

Na região NIR, as ocorrências espectrais correspondem aos sinais de absorção, transmitância e também reflectância, que estão relacionados aos sobretons e combinações de transições fundamentais que ocorrem na região do MIR. As ligações envolvidas nas transições vibracionais ativas no NIR são tipicamente C–R, N–H e O–H presentes majoritariamente nas estruturas de carboidratos, lipídios, ácidos nucleicos, proteínas dentre outras estruturas biologicamente ativas nos organismos vivos, o que torna sua aplicação nas ciências biológicas bastante interessantes e ricas de informações. O alcance da radiação infravermelha ocorre em todas as direções da superfície da amostra em contato com a sonda de fibra ótica responsável por propagar o fluxo de energia da fonte à amostra. O

processo de dispersão é predominante quando os materiais são fracamente absorventes e quando a penetração da radiação é grande em relação ao comprimento de onda incidente, facilitando a obtenção de sinais a partir de amostras sólidas sem necessidade de alterações físicas ou tratamento químico do material.

Os mecanismos de medida na região do infravermelho próximo se classificam de acordo com o tipo de interação da radiação com a amostra, sendo os principais o modo de transmitância, reflectância e reflectância difusa. Os espectros, embora semelhantes na aparência, não são idênticos aos espectros de absorção correspondentes, pois carregam tanto informação química quanto física. O mais utilizado para análise de amostras sólidas é o modo de reflectância difusa, onde a radiação reflete de forma aleatória ao entrar em contato com a amostra, apresentando eficiência na obtenção de dados espectrais capazes de fornecer informações qualitativas e quantitativas a cerca da amostra, tornando-se uma ferramenta analítica muito útil para investigação de diversos tipos de sólidos, principalmente tecidos biológicos, como mostrada na Figura 3.

Figura 3: Representação do espectrômetro FT-NIR (A) e obtenção de espectro por Reflectância difusa em crânio de rato (B).



Fonte: Autor

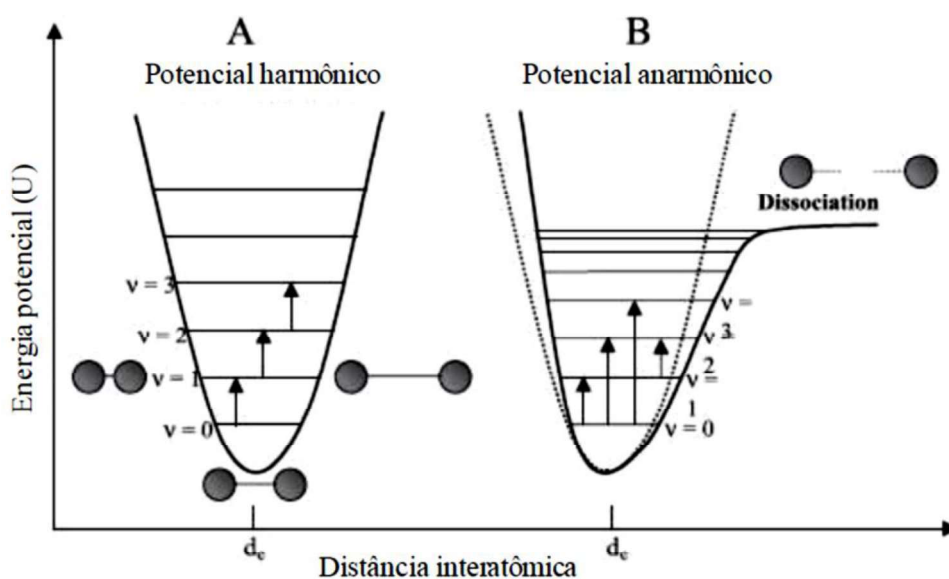
Naturalmente, os átomos envolvidos em ligações químicas apresentam um movimento espontâneo, ou seja, eles não estão fixos e estáticos, uma vez que estão continuamente desenvolvendo movimentos vibracionais e rotacionais em torno de um eixo, ou átomo central que pode ser relacionado de forma proporcional à energia envolvida no processo. Basicamente, tais movimentos podem ser classificados como estiramentos ou deformações angulares, podendo, ambos, serem simétricos ou assimétricos [16]. Os estiramentos são vibrações que ocorrem ao longo do eixo da ligação química entre átomos,

e por consequência ocorre o alongamento ou encurtamento da ligação. Esse movimento é considerado simétrico quando os átomos se afastam e se aproximam do átomo central ao mesmo tempo, ou assimétrico, quando à medida que um átomo vibrante se afasta, o outro se aproxima do átomo central. Por outro lado, as vibrações por deformação angular consistem em movimentos que um átomo realiza e alteram o ângulo de ligação entre três átomos.

No clássico modelo do oscilador harmônico, a energia potencial (U) de uma vibração será função do afastamento dos átomos. Apesar de útil para o entendimento do conceito de energia vibracional, esse modelo apresenta uma limitação quando são considerados sistemas microscópicos, como no caso das moléculas, pelo fato de que tais sistemas não assumem perfis contínuos de energia, como seria previsto por esse modelo clássico. De acordo com a mecânica quântica, tais sistemas moleculares só podem assumir níveis discretos de energia [15,16].

De acordo com esse modelo quântico/harmônico, as transições entre os diferentes níveis vibracionais adjacentes só podem acontecer quando $\Delta v = \pm 1$. Essa diferença de energia entre os níveis é sempre a mesma. Para que uma molécula absorva energia e, conseqüentemente, seja promovida até um nível vibracional excitado, a radiação incidente deve corresponder exatamente à diferença entre os dois níveis energéticos adjacentes [16].

Figura 4 – Representação esquemática dos modelos Harmônico (A) e Anarmônico (B) para a energia potencial de uma molécula diatômica.



Fonte: Referência [12]

Podemos ainda observar outro fator importante do modelo harmônico, para esse modelo as vibrações são independentes entre si, dessa forma as bandas de combinação NIR não poderiam existir, portanto não deveriam ser observadas [12]. Todavia, tanto as bandas de combinação quanto os sobretons existem de fato e são experimentalmente visualizados. A anarmonicidade também pode estar presente nas propriedades elétricas das moléculas. Especificamente, isso irá afetar no momento de dipolo, que em um modelo anarmônico, não tem dependência linear com a distância interatômica. Esse tipo de anarmonicidade pode fornecer caminhos para ocorrência de sobretons e bandas de combinação, mesmo se nenhum desvio mecânico do modelo harmônico for observado no sistema em questão, tornando os espectros NIR ainda mais complexos.

A literatura atual demonstra a evolução e versatilidade da espectroscopia NIR quando aplicada na resolução de inúmeros problemas da indústria e centros de pesquisa. Estudos em sistemas aquáticos têm sido reportados explorando mudanças espectrais relacionadas aos rearranjos moleculares da água causados por pequenas mudanças nos solutos dos sistemas aquosos, fornecendo uma alternativa rápida e precisa para medidas analíticas clássicas de sacarídeos, mesmo em níveis de concentração milimolar [17]. O combate à adulteração de alimentos também tem usufruído do potencial do FT-NIRS combinado com métodos multivariados no qual foi desenvolvida uma nova ferramenta alternativa para a determinação rápida da adulteração de leite de vaca em amostras de leite de camelo [18]. As ciências forenses tem sido uma grande demandante da evolução dessa ferramenta analítica, em que novas ferramentas de triagem são cada vez mais necessárias para a detecção precoce de drogas ilícitas em produtos falsificados [19].

A espectroscopia de infravermelho próximo, seguida de quimiometria, demonstrou ser uma ferramenta útil para a caracterização multiparamétrica de matrizes complexas. Na medicina, também chamada técnicas foto-médicas, o uso de fontes de excitação de infravermelho próximo (NIR) em nanopartículas ativada por NIR pode oferecer penetração de luz mais profunda melhorando os tratamentos e a biodetecção dentro de volumes de tecido mais altos e com menos efeitos colaterais (por exemplo, superaquecimento) podem ser implementados com sucesso no tratamento de tumores cancerígenos [20].

Como todas as metodologias, a espectroscopia NIR apresenta vantagens e desvantagens do ponto de vista analítico. Os aspectos positivos ficam por conta de ser uma técnica não destrutiva, o tratamento da amostra é mínimo, diminuindo assim a presença de interferentes químicos, quando comparada com outras técnicas possui rápida aquisição

espectral minimizando tempo de análise, pode ser considerada uma técnica não invasiva com custo relativamente baixo, pois não há necessidade de reagentes. A complexidade do espectro NIR, a baixa seletividade na detecção de componentes minoritários, a dependência de ferramentas computacionais para retirada de ruídos indesejáveis e correção de linha base são algumas limitações instrumentais dessa ferramenta.

A tecnologia NIR tem sido há mais de quatro décadas, capaz de incorporar todos os benefícios trazidos pela evolução de campos correlatos, como Quimiometria, novos materiais para componentes ópticos, novos sensores e matrizes de sensores, microcomputadores e microeletrônicos. Sem dúvida, o NIR apresenta alta capacidade de incorporar rapidamente os avanços nos campos correlacionados garantindo um campo extenso e frutífero para pesquisa e desenvolvimento [11].

2.1.2 Espectroscopia no infravermelho médio (MIR)

A espectroscopia no infravermelho médio (MIR) é uma técnica vibracional amplamente conhecida e explorada em nível de pesquisa científica quanto em processos industriais. A espectroscopia estuda o comportamento de determinados compostos frente à radiação [21], e sua aplicação tem sido avaliada como uma forte tecnologia para pesquisas na área biológica [22,23], onde é conhecida como bioespectroscopia. A faixa do infravermelho médio (MIR) apresenta uma melhor resolução quando comparada a outras regiões do espectro, uma vez que pode ser diretamente relacionada com grupos funcionais moleculares [24].

A região do MIR compreende a faixa entre 400 cm^{-1} e 4000 cm^{-1} do espectro eletromagnético e essa radiação são absorvidos por biomoléculas [25,26]. Por ser uma técnica fortemente versátil, diversas informações químicas valiosas acerca de composição, estrutura e conformação de moléculas podem ser extraída da composição espectral. Cada molécula apresenta um espectro único e característico, em função de sua composição e estrutura química.

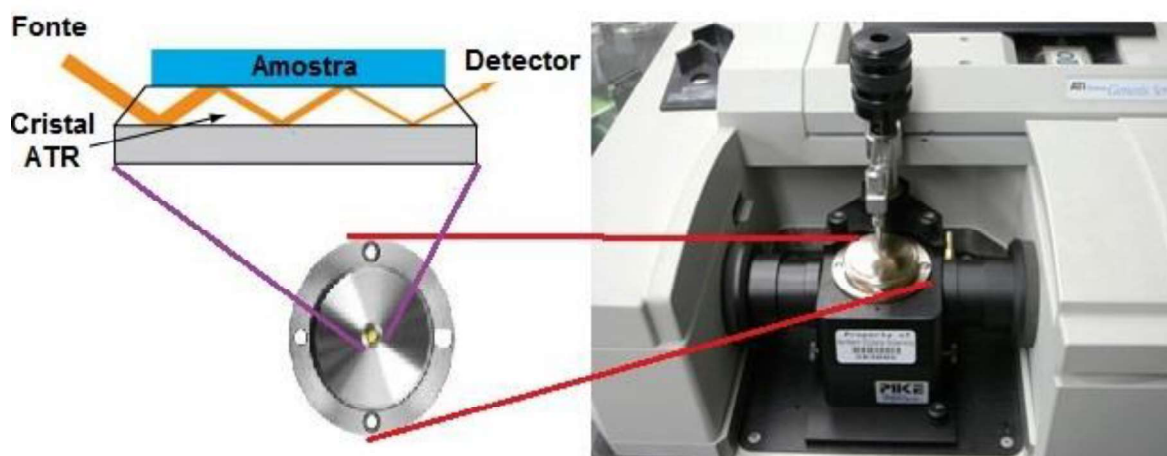
Além de análises qualitativas, a espectroscopia no infravermelho também permite que medidas quantitativas sejam realizadas, levando em consideração a fração da radiação incidente que foi absorvida pela amostra, através da bem estabelecida lei de Beer [23,24]. Atualmente, os espectrofotômetros de infravermelho são, em sua grande maioria, equipados com um sistema óptico (interferômetro de Michelson) que, associado a

transformações matemáticas conhecidas como Transformada de Fourier (FT), permitem de forma rápida a obtenção de espectros com alta relação sinal/ruído e alta resolução.

A reflectância total atenuada (ATR) associada à espectroscopia no infravermelho por transformada de Fourier, ATR-FTIR, vem crescendo consideravelmente nas principais investigações na região do infravermelho médio. Isso se deve ao fato de que para amostras sólidas e líquidas, incluindo tecidos, células, fluidos biológicos, insetos no qual o modo ATR contorna os diversos inconvenientes das análises por infravermelho, no qual podemos citar a preparação de amostras e a reprodutibilidade espectral.

O modo de aquisição por ATR permite obter espectros de forma estável, robusta, não destrutiva e com mínima ou nenhuma preparação de amostras, uma vez que estas podem ser posicionadas diretamente sobre o cristal de ATR, estando a amostra viva ou não e então a medida é realizada. A reflectância total é um caso especial de reflexão de uma onda eletromagnética em uma interface entre dois meios.

Figura 5: Representação do espectrômetro ATR FT-IR destacando a interação entre a radiação e a amostra posicionada no cristal de reflectância total atenuada.



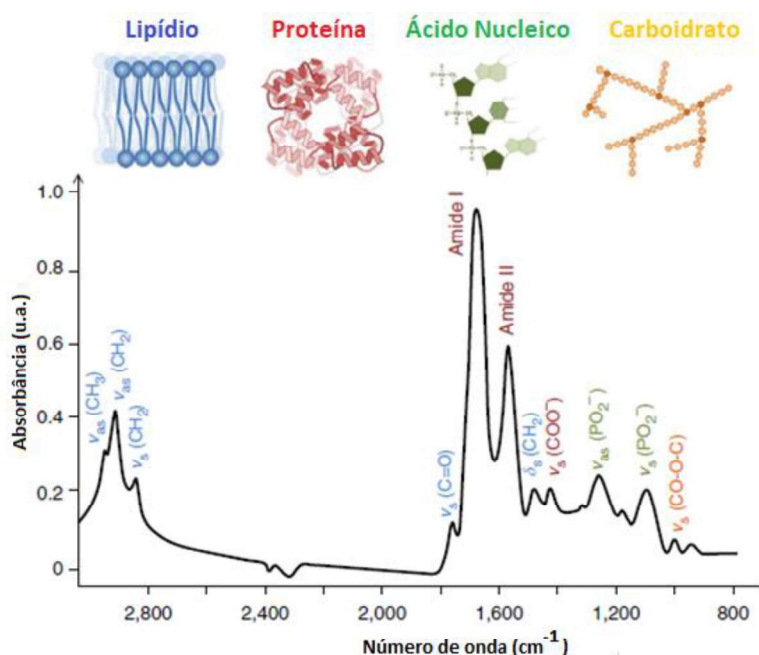
Fonte: Autor

O princípio da espectroscopia ATR-FTIR ocorre através da incidência da radiação infravermelha sob um ângulo específico e um meio de reflectância interna opticamente densa (frequentemente um “cristal” de diamante) que reflete a radiação na interface com um meio mais denso opticamente, criando uma onda evanescente que se estende para além da superfície do cristal e pode interagir com as amostras colocadas na interface, permitindo, então, medidas de absorção.

Essa onda evanescente se sobressai apenas alguns microns ($0.5 - 5 \mu$) transpassando o cristal, portanto, a amostra deve estar perfeitamente fixa na superfície do cristal. Nas regiões do infravermelho onde a amostra absorve energia, a onda evanescente será atenuada ou alterada, e isto será medido pelo detector, gerando um espectro de infravermelho.

A faixa do espectro do MIR entre 900 cm^{-1} a 1800 cm^{-1} é conhecida como impressão digital biológica [23], pois é nesta faixa onde pode-se encontrar informações referentes a biomoléculas como lipídeos (1750 cm^{-1}), carboidratos (1155 cm^{-1}), proteínas (amida I - 1650 cm^{-1} , amida II, 1550 cm^{-1} , amida III, 1260 cm^{-1}), DNA/RNA (1225 cm^{-1} , 1080 cm^{-1}) [23–25]. A espectroscopia do infravermelho com reflexão total atenuada utilizando transformada de Fourier (ATR-FTIR) foi identificada como uma forma de se coletar e analisar eficientemente espectros de amostras biológicas [23,21] utilizando informações vibracionais de biomoléculas como demonstrada na Figura 6.

Figura 6: Representação da região de impressão digital biológico destacando as macromoléculas de maior importância biológica e suas respectivas regiões de absorção no infravermelho médio.



Fonte: Adaptado referência [27]

A biospectroscopia através da utilização do FTIR tem permitido a identificação eficiente de biomoléculas [28], a análise quantitativa de proteínas [29], identificação e diferenciação de infecções virais [30,21], fúngicas [31] e bacterianas [32,33].

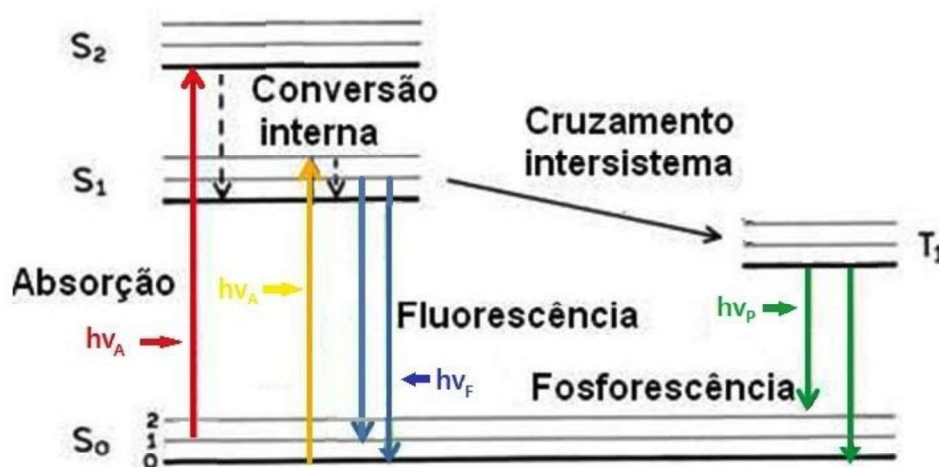
2.1.3 Fluorescência Molecular

A espectroscopia de fluorescência molecular se baseia na propriedade de emissão de luz que certas moléculas apresentam quando são excitadas por uma radiação luminosa incidente. Essas moléculas capazes de fluorescer são chamadas de fluoróforos, e podem ser detectadas, em mínima quantidade, contra um fundo preto (com ausência de luz de fundo) e uma relação alto sinal e baixo ruído otimizada [34].

A fluorescência é um processo consideravelmente seletivo, pois para alcançar o estado de fluorescência alguns requisitos moleculares devem ser contemplados, aumentando a probabilidade de haver emissão pela molécula [35]. Moléculas aromáticas, com alta rigidez, planares e que contêm duplas ligações conjugadas, possuem a habilidade de emitir radiação na forma de luz, um fenômeno conhecido como luminescência. Para que a luminescência ocorra às moléculas devem estar em seu estado fundamental (de mais baixa energia) então com a absorção de energia elas vão para seu estado eletronicamente excitado [27]. Após a absorção, a radiação na região do ultravioleta (acima de 200 nm) e visível do espectro eletromagnético leva a transições eletrônicas entre os elétrons não ligantes (elétrons “n”) e elétrons de ligações π do tipo $n \rightarrow \pi^*$ e $\pi \rightarrow \pi^*$. Posteriormente, as moléculas excitadas podem retornar ao seu estado fundamental emitindo parte da energia (regra de Stokes) que foi absorvida na forma de luz [36].

Dependendo da natureza do estado excitado, podemos dividir a fotoluminescência em fluorescência e fosforescência. Se no estado excitado o elétron envolvido na transição mantém a sua orientação de spin no orbital excitado, continuando emparelhado com o segundo elétron que está no orbital do estado fundamental, tem-se o estado excitado singlete, o que caracteriza a fluorescência. Já se o elétron no estado excitado inverte a sua orientação de spin, tem-se o estado excitado tripleto que caracteriza a fosforescência. O diagrama de Jablonski (Figura 7) demonstra de forma simples os possíveis processos: absorção de fótons, conversão interna, fluorescência, cruzamento inter sistema, fosforescência, fluorescência com retardo e transições tripleto-tripletto [36,37].

Figura 7: Diagrama de Jablonsky representando os estados possíveis de transição eletrônica.



Fonte: Adaptado da referência [35]

Nesse diagrama, processos de desativação podem ser observados, são eles:

- Conversão interna – é uma transição não-radiativa entre dois estados eletrônicos de mesma multiplicidade de spin. Em solução, este processo é seguido por uma relaxação vibracional para o nível vibracional mais baixo do estado eletrônico final. O excesso de energia vibracional pode inclusive ser transferido para o solvente durante colisões da molécula excitada com as moléculas do solvente na sua vizinhança [27].
- Fluorescência – a emissão de fótons que acompanha a relaxação S₁ → S₀ é chamada de fluorescência. Deve ser enfatizado que, apesar de algumas exceções, a emissão de fluorescência ocorre a partir de S₁ e suas características (exceto polarização) não dependem do comprimento de onda de excitação (considerando que somente exista uma espécie no estado fundamental). O espectro de fluorescência é localizado em comprimentos de onda maiores (energia mais baixa) que o espectro de absorção, por causa da perda de energia no estado excitado devido à relaxação vibracional [27,36]. De acordo com a regra de Stokes, o comprimento de onda de emissão de fluorescência deve ser sempre maior que o da absorção. Entretanto, na maioria dos casos, o espectro de absorção intercepta parcialmente o espectro de fluorescência, isto é, uma fração de luz é emitida em comprimento de onda menor que a luz absorvida. Tal observação parece estar, em um primeiro momento, em contradição com o princípio da conservação da energia. No entanto, isto pode ser explicado pelo fato de que em temperatura ambiente uma pequena fração de moléculas está em um nível vibracional maior que o nível 0 tanto no estado fundamental como no estado excitado.

- Cruzamento intersistema – é uma desativação não radiativa entre dois níveis vibracionais isoenergéticos pertencentes a estados eletrônicos de multiplicidades diferentes. Por exemplo, uma molécula excitada no nível vibracional 0 do estado S1 pode mover-se para o nível vibracional isoenergético do estado tripleto.

- Tripleto-tripletto – Quando uma molécula está excitada e alcança o estado tripleto T1, ela pode absorver outro fóton em um comprimento de onda diferente, porque transições tripleto-tripletto são permitidas pelo spin. Estas transições podem ser observadas quando a população de moléculas no estado tripleto é grande o suficiente, o que pode ser conseguido pela iluminação com um pulso intenso de luz [27,36,37].

O instrumento capaz de gerar um espectro de fluorescência é o espectrofluorímetro. Nesses instrumentos a fonte de luz é geralmente uma lâmpada de xenônio de alta pressão, que oferece a vantagem de emissão contínua do ultravioleta até o infravermelho. Dois monocromadores são usados para selecionar comprimentos de onda de excitação. Um dos monocromadores é direcionada em um ângulo reto em relação ao feixe incidente de modo a capturar a fluorescência emitida nessa direção. Varreduras automáticas dos comprimentos de onda são realizadas por monocromadores motorizados, que são controlados por dispositivos eletrônicos e pelo computador, no qual os dados são armazenados. Esse sistema de aquisição de dados facilitou a obtenção de uma matriz de excitação-emissão (EEM), em que se obtêm os espectros de excitação e de emissão de uma amostra. Esses espectros são produzidos pela variação do comprimento de onda de emissão e varrendo-se um intervalo de excitação [35]. Alguns fatores influenciam diretamente a qualidade da matriz de excitação-emissão obtida, sendo a concentração uma das principais. É recomendado que as amostras analisadas suficientemente diluídas (apresentem absorvância de até 0.05), principalmente nos casos de compostos altamente fluorescentes, possibilitando uma relação linear entre a concentração e o sinal obtido. Outros fenômenos como o efeito do filtro interno (*inner filter effect*), a supressão de fluorescência (*quenching*), a presença de espalhamento Rayleigh e bandas de difração de segunda ordem geradas pela grade de difração ocorrem com frequência podendo levar a medidas incorretas [38].

Avanços na espectroscopia de fluorescência tem fornecido uma alternativa às abordagens tradicionais para caracterizar a matéria orgânica dissolvida em ambientes aquáticos, permitindo uma caracterização rápida e precisa [39]. Em sistemas biológicos, a

observação de múltiplos emissores de fluorescência é comum, possibilitando a aplicação dessa ferramenta de maneira muito efetiva em diversas áreas. A evolução em hardware, software e marcadores fluorescentes continua a abrir novas perspectivas para aplicações. Exemplos como controle de qualidade das proteínas do ovo [40], na investigação da hidratação de regiões de dehalogenases funcional com base na incorporação *in vivo* de aminoácidos não naturais fluorescentes e sua análise por espectroscopia de fluorescência em estado estacionário [41]. A sensibilidade dessa técnica permite até mesmo visualizar alterações a nível celular capazes de desmascarar doenças e desenvolver métodos efetivos para o rastreamento do câncer [42].

2.1.4 Voltametria de Pulso Diferencial

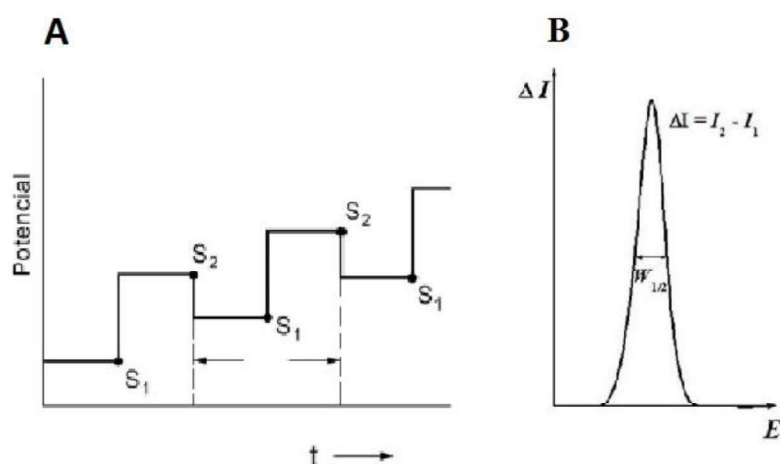
As técnicas voltamétricas correspondem a um conjunto de métodos eletroanalíticos capazes de obter informações sobre um determinado analito a partir da intensidade de corrente obtida sob um potencial aplicado. Esta técnica usa basicamente três eletrodos sendo um eletrodo de trabalho, um eletrodo de referência e o eletrodo auxiliar. As medições das intensidades de corrente são realizadas entre os eletrodos de trabalho e auxiliar e o potencial aplicado é medido entre os eletrodos de trabalho e de referência. O uso de eletrólitos de suporte é necessário, no qual consiste em um sal de metal, que reduz os efeitos de migração de carga e diminui a resistência da solução [43].

Durante a eletrólise, nos processos redox que ocorrem na interface eletrodo-solução acontece transferência de massa (migração de espécie da solução para a interface eletrodo-superfície) e transferência de carga (transferência de elétrons na superfície do eletrodo, que ocorre numa região de interface eletrodo-solução, chamada de dupla camada elétrica) [43,44]. O transporte de massa pode ocorrer por migração, convecção ou por difusão. No processo de migração o movimento de íons através da solução é causado pela atração ou repulsão entre as espécies iônicas em solução e o eletrodo de trabalho. Na convecção ocorre a movimentação das espécies causadas por perturbação mecânica do fluxo da solução. E por difusão há movimentação espontânea das espécies química devido à formação de um gradiente de concentração do analito. Os processos de migração são minimizados quando utilizamos um excesso de eletrólito suporte capaz de impedir a formação de um gradiente de cargas. Desse modo, o transporte de massa em voltametria é realizado basicamente por difusão [43,44].

As técnicas de voltametria de pulso são amplamente adotadas em eletroanálise objetivando determinar espécies eletroativas, alcançando elevadas sensibilidades, particularmente na presença de oxigênio dissolvido que podem gerar correntes com ruídos de fundo. Nesta técnica, a corrente capacitiva é proporcional a $e^{-t/RC}$, onde t é o tempo, R a resistência da solução e C a capacitância da dupla camada elétrica e a corrente faradaica é aproximadamente proporcional a $t^{1/2}$. Após a aplicação do pulso de potencial, o decaimento da corrente capacitiva é muito mais rápido que o da corrente faradaica, assim as medidas de corrente somente são realizadas após a contribuição da corrente capacitiva ter sido minimizada [45]. Por causa do seu desempenho, as técnicas modernas de pulso têm substituído à polarografia ou voltametria clássica na rotina dos laboratórios de química analítica.

A voltametria de pulso diferencial (DPV) é uma técnica extremamente útil para determinar traços de matéria orgânica e inorgânica em solução. Na PDV, pulsos de potencial de igual amplitude são aplicados sobre uma rampa linear ou escada de potencial em intervalos de tempo e a corrente I_1 é medida antes do pulso ser aplicado (S_1) e logo no final do pulso (S_2) é medido a corrente I_2 , como mostra a Figura 8A. O tempo t durante o qual o pulso de potencial é aplicado define a largura do pulso. Estas correntes são computacionalmente subtraídas, já que a primeira é principalmente a contribuição da corrente capacitiva e a segunda a contribuição da corrente faradaica, e então registradas contra o potencial, gerando o voltamograma de pulso diferencial ilustrado na Figura 8B.

Figura 8: Representação do processo de (A) Excitação da DPV no início do pulso (S_1) e no final do pulso (S_2) e (B) Voltamograma de pulso diferencial.



Fonte: Autor

A amplitude escolhida para o pulso de potencial deve levar em consideração a reação em estudo e a sensibilidade exigida, uma vez que, grandes amplitudes geram correntes capacitivas residuais que podem diminuir a possibilidade de detectar a corrente faradaica. Para pequenas amplitudes de pulso, a largura da meia altura ($W_{1/2}$) do pico de corrente é inversamente proporcional ao número de elétrons transferidos ($90/n$), o que permite caracterizar o mecanismo da reação de transferência de carga. A corrente de pico (I_p) é diretamente proporcional à concentração do analito, mas também depende da velocidade da reação eletroquímica [46,47].

A PDV tem sido amplamente explorada quando há necessidade de identificar ou monitorar a presença de um analito de interesse. Tem sido desenvolvidos métodos para monitoramento rápido, em tempo real, de Ni^{+2} em água de mina [48], um método de controle de qualidade para determinar quantitativa e qualitativamente a eficiência de remoção de tocoferóis do óleo de peixe comercial [49], na quantificação do ácido trans-retinóico presente na vitamina A e administrado por via parenteral para fins de terapia do câncer e foi aplicado para determinar capsacinóides em pimenta e óleos de oliveira [50]. O procedimento proposto é rápido, simples e permite o monitoramento de amostras complexas em tempo real. No entanto, em sistemas biológicos complexos como estruturas celulares, insetos, parasitas, plasma sanguíneo e tecido essa metodologia ainda apresenta um grande caminho de exploração.

2.2 Análise Quimiometria

2.2.1 Análise por Componentes Principais (PCA)

A PCA é o método de análise exploratória de dados mais utilizado na quimiometria. Esse método permite a redução de dimensionalidade de um grande conjunto de dados em um novo sistema, chamado componente principal (PC), o que possibilita a visualização da natureza multivariada dos dados em dimensões inferiores. As PC's são as representações dos novos eixos formados a partir de combinações lineares das variáveis originais e nelas estão contidas as informações mais importantes que descrevem os dados. Elas são sempre ortogonais umas as outras para que a variabilidade contida em uma PC não permaneça nas demais, então sucessivas PC's descrevem quantidades decrescentes da variância explicada dos dados.

Matematicamente, uma matriz de dados X é decomposta no produto de duas outras matrizes, denominadas *scores*(T) e *loadings* (P), somadas a uma matriz resíduos(E), como mostrado na equação 1.

$$X = T.P^t + E \quad \text{Eq. 1}$$

Os scores correspondem a projeção das amostras em um vetor de dados formado pelos PC's, indicando a similaridade entre as amostras e os loadings corresponde ao peso que cada variável tem no conjunto de dados, demonstrando, portanto, a contribuição de cada variável na determinação de um PC. Contudo, os resíduos tem a função de regular o produto $T.P$, sendo assim quando essa matriz resíduos tende a zero a variabilidade do sistema é explicada ao máximo. Nesta tese, a PCA foi utilizada para fins de análise supervisionada com o auxílio de um classificador por análise discriminante linear (LDA) e análise discriminante quadrático, de modo a se observar a melhor classificação das classes de amostras analisadas.

2.2.2 Seleção de variáveis – Algoritmo de Projeções Sucessivas (SPA) e Algoritmo Genético (GA)

Em análise multivariada é muito comum problemas associado ao tamanho da matriz, pois quanto mais variáveis um dados apresenta maior será a dificuldade de extrair informações pertinentes ao entendimento do mesmo. Em sistemas biológicos, conjuntos de dados com poucas amostras e/ou um elevado número de variáveis agregam ao sistema uma dificuldade a mais para sua completa interpretação. Muitas dessas variáveis são irrelevantes ou redundantes, e as que são importantes são frequentemente desconhecidas ou possuem contribuição duvidosa. Algoritmos de seleção de variáveis surgem, portanto, como uma alternativa valiosa para minimizar ou contornar problemas desse tipo.

O GA é um método de seleção de variáveis de natureza estocástica [51]. Contudo, o procedimento é realizado através de uma simulação de processos naturais da evolução com aplicação da teoria da evolução das espécies proposta por Darwin. O GA seleciona as variáveis por etapas, avaliando o conjunto que melhor se adapta a uma possível solução. Inicialmente há seleção das variáveis, em seguida há o cruzamento e posterior mutação, portanto as variáveis com melhor aptidão baseada no menor valor de uma função de custo constituem um ciclo de evolução e tornam-se as variáveis selecionadas que melhor classificam os dados.

O SPA foi inicialmente desenvolvido para ser aplicado em modelos de calibração multivariada, com o intuito de resolver problemas de multicolinearidade através de uma

série de projeções ortogonais [52]. O SPA mostrou ser uma técnica eficiente para seleção de variáveis espectrais no contexto da calibração multivariada, especificamente quando aplicado à regressão linear múltipla. Quando aplicado a problemas de classificação, o SPA escolhe as variáveis com menor multicolinearidade e que levam a menor valor de classificação incorreta para o modelo. É importante ressaltar que em métodos de classificação, a capacidade de classificação da análise discriminante linear (LDA) pode ser comprometida pela presença de multicolinearidade entre as variáveis. Por essa razão, a LDA se restringe normalmente a problemas de pequenas dimensões. Dessa forma, a minimização de colinearidade proporcionada pelo SPA deve ser útil também para modelos LDA.

2.2.3 Análise discriminante

Os classificadores mais conhecidos e explorados são LDA, análise discriminante quadrática (QDA). LDA é um dos algoritmos mais comuns utilizados na classificação supervisionada, enquanto o QDA raramente é usado devido à escassez de algoritmos para esse método. Tanto LDA quanto QDA calculam a função discriminante entre duas classes de acordo com (1) uma métrica de distância de Mahalanobis; (2) um termo de covariância, que em LDA é a matriz de covariância agrupada e na QDA é a matriz de variância-covariância de cada classe; e (3) um termo de probabilidade. Embora a LDA mostre bons resultados de classificação a alta variabilidade de alguns sistemas biológicos complexos, seu desempenho pode diminuir devido à alta variabilidade intra-classe. Isso ocorre porque a LDA usa algumas aproximações, uma vez que as classes são consideradas como tendo matrizes de covariância e probabilidades de distribuição por similaridade [53]. O QDA é uma boa opção para lidar com classes com diferentes estruturas de variância, uma vez que calcula a matriz de variação-covariância de cada classe individualmente como regra de classificação.

A análise por componentes principais e algoritmos de seleção de variáveis tais como SPA e GA são ótimas alternativas para a redução da dimensionalidade dos dados e, nesse caso, resolver os problemas de multicolinearidade frequentemente encontrada por sistemas em que o número de amostras é muito menor que o número de variáveis dos dados.

2.2.4 Algoritmo 2D

A matriz de excitação-emissão (EEM) de fluorescência molecular é considerada um dado químico de segunda ordem, pois forma um tensor de dados. Existem poucos algoritmos de classificação para dados de segunda ordem quando comparados com os métodos descritos nessa tese. As formas mais utilizadas de lidar com dados de segunda ordem são o uso do desdobramento, procedimentos ou algoritmos que reduzem a dimensionalidade dos dados, como a análise de fatores paralelos (PARAFAC) [54,55]. Ao fazer isso, a estrutura de um tensor é transformada em uma matriz, semelhante a um espectro de infravermelho, e então os algoritmos de classificação são aplicados.

A proposta do algoritmo 2D, diferentemente do PARAFAC, é não desdobrar o tensor de dados, aplicando os classificadores diretamente em um dados tridimensional. Primeiramente aplica-se a PCA nos dados e uma matriz de *score* é gerada e em seguida cada uma é organizada em um corpo tridimensional. Em seguida, a matriz de *scores* é utilizada como o dado de entrada para a classificação supervisionada.

3 Objetivos

3.1 Objetivo Geral

A presente tese de doutorado tem por objetivos avaliar o desempenho de métodos instrumentais de análise em conjunto classificação multivariada aplicada a diferentes matrizes biológicas.

3.2 Objetivos Específicos

- Observar o desenvolvimento molecular do reparo ósseo utilizando NIR e classificação multivariada, avaliando a eficiência dos tratamentos em comparação com o crânio saudável.
- Desenvolver um método de triagem de idosos com e sem sarcopenia a partir de plasma sanguíneo, utilizando ATR FTIR e classificação multivariada.
- Desenvolvimento de um método usando DPV e espectroscopia de fluorescência combinada com análise multivariada para discriminar grupo controle e com a presença de flunitrazepam em larvas de moscas necrófagas.
- Comparar a eficiência do algoritmo 2D-LDA em relação ao PARAFAC-LDA.

4 Metodologias

Os estudos reportados nos Capítulos 2, 3, 4 e 5 desta tese foram realizados pela colaboração entre o Instituto de Química (IQ/UFRN) e os Departamentos de Biofísica e Farmacologia (DBF/UFRN), Departamento de Microbiologia e Parasitologia da UFRN, Departamento de fisioterapia da UFRN e Departamento de fisioterapia da Universidade Federal de Alfenas (UNIFAL-MG), Minas Gerais, Brasil. Também se realizou colaboração com o Departamento de Química da Universidade Federal da Paraíba (DQ/UFPB), Paraíba, Brasil. Os animais mortos para realização dos experimentos foram aprovados pelo comitê de ética (CEUA) da Universidade Federal do Rio Grande do Norte (UFRN), número do protocolo #067/2014 e #044/2013 e o Comitê Institucional de Ética em Pesquisa da Universidade Federal de Alfenas (UNIFAL-MG), #564.941 e realizado de acordo com a Declaração de Helsinque.

4.1 Coleta e preparação das amostras

4.1.1 Ratos utilizados para avaliar defeito crítico

Foram utilizados 45 ratos Wistar machos adultos (de aproximadamente 6 meses), pesando em média 320gramas. Foram divididos em um grupo controle sem defeito craniano contendo 5 animais e 4 grupos com defeito craniano, cada um com 10 animais, onde em cada um desses animais foi criado um defeitos cranianos simétricos de 5mm.

4.1.2 Formação do defeito

Os animais foram anestesiados com uma injeção intraperitoneal de cloridrato de cetamina (10%), na dosagem de 80mg/kg, e Xilazina (2%) na dosagem de 10mg/kg. Uma incisão na pele feita a partir da área do nariz frontal na protuberância do osso occipital externo, expondo a extensão da calvária. Um defeito craniano simétrico de espessura total de 5 mm foi feito entre as duas regiões parietais, no centro do crânio, com uma broca de trefina acionada por um micromotor elétrico de 30.000 rpm (NSK®), sob irrigação constante de solução fisiológica estéril para evitar superaquecimento do osso.

4.1.3 Tratamento por laser

Os grupos que receberam a incidência de laser diodo (Kondortech – BioWave LLLT Dual,Brazil) na dosagem de 120J/cm² (potência de 50 mW, tempo de exposição de 10 s) e comprimento de onda de 780nm, com aplicações feitas em 4 pontos equidistantes

em torno do defeito com distância de 1mm da borda, logo após o procedimento cirúrgico, 48 horas e 96 horas depois, protocolo adaptado do artigo de Marques et al.,(2015). A esponja de colágeno utilizada era da marca Hemospon® (Tecnew, 10 Und.), ela foi cortada com uma lâmina de bisturi no tamanho aproximado do defeito, com espessura de aproximadamente 1mm.

Foram implantados os arcabouços nos grupos experimentais (G1 e G3) e um grupo recebeu a incidência do laser (G3), enquanto o outro não (G1). A primeira incidência foi logo após o procedimento cirúrgico, e foram feitas irradiações a cada 48 horas por 5 dias. Já o Grupo G2 recebeu apenas a incidência do laser. Para cuidado pós-operatório foi utilizado analgésico Dipirona monoidratada 500mg, uma gota via oral (gavagem) durante 5 dias, e 5 gotas no bebedouro por 1 semana.

4.1.4 Plasma Humano

Antes da realização das coletas todos os participantes forneceram consentimento informado por escrito antes de participar deste estudo. A coleta das amostras de sangue foi realizada em tubos à vácuo, com anticoagulante EDTA. As amostras foram centrifugadas em 4000 rpm, por 10 minutos, e o plasma removido em ambiente estéril, dentro de 30 minutos após a coleta, e as alíquotas estocadas em freezer -80°C até as análises.

4.1.5 Ratas utilizadas para identificação de flunitrazepam

Para realizar as experiências, foram utilizadas 32 ratas Wistar (*Rattus Norvegicus*), com peso médio de 255 ± 30 g (Protocolo de Aprovação do Comitê de Ética 044/2013). As ratas com idade de 60 ± 5 dias utilizados no início do experimento do Biotério do Departamento de Biofísica e Farmacologia da Universidade Federal do Rio Grande do Norte foram acomodados (4 animais / gaiola) em gaiolas/viveiro com acesso gratuito a soluções de água (ou etanol) e comida (ração Purina®, Labina). As ratas foram divididas em quatro grupos com 8 ratas cada, em que cada grupo recebeu tratamento diferente: (1) etanol - em que o álcool etílico (Alcoolabor®, Segmenta) foi utilizado em concentrações crescentes de 2% (3 dias), 4% (3 dias), 6% (15 dias) totalizando um período de 21 dias [56]; (2) flunitrazepam - dosagem de 2 mg / kg oralmente [57]; (3) conjugado (etanol e flunitrazepam) - álcool etílico (Alcoolabor®, Segmenta) foi utilizado em concentrações crescentes de 2% (3 dias), 4% (3 dias), 6% (15 dias) totalizando um período de 21 dias [56] e recebeu doses de 2 mg / kg de flunitrazepam por via oral [57]; (4) controle - recebeu

apenas água ad libitum. Após uma hora de administração de flunitrazepam, as ratas foram eutanasiadas por decapitação com guilhotina e cada rata foi colocada em sacos de plástico individualizados. Os animais pertencentes aos quatro tratamentos foram transportados separadamente.

No campo, as armadilhas de cheiro contendo os ratos foram distribuídas em uma 8 *grids* (distância da grade de aproximadamente 50 m e a 1,50 m de altura do solo) ao longo de uma das trilhas do 7º Batalhão de Engenharia de Combate (Exército Brasileiro), uma área da floresta atlântica localizada na cidade de Natal / Brasil. Cada *grid* continha uma rata de cada tratamento (controle, etanol, flunitrazepam e conjugado) separados por 5 m. As armadilhas permaneceram no campo por cinco dias consecutivos. No terceiro e quarto dias, foram coletadas 10 imaturos (resultantes da oviposição e larviposição, priorizando as maiores larvas, representando as primeiras gerações que chegaram na carcaça) de cada armadilha. As larvas foram colocadas em um frasco de vidro contendo glicerina e enviadas ao laboratório para identificação e análise química.

Em laboratório, um conjunto de 3 larvas foi adicionada a 2 mL de dimetilsulfóxido (DMSO), depois triturado, homogeneizado e centrifugado. A solução sobrenadante foi armazenada em criotubos e depois enviada para análise. Foram formados dois grupos para classificação, (1) controle mais etanol - larvas coletadas em carcaças dos grupos de controle e etanol; (2) flunitrazepam - larvas coletadas em grupos flunitrazepam e conjugado. No total, 40 amostras (20 controle e 20 contendo flunitrazepam) foram direcionadas para DPV e análise de fluorescência molecular.

4.2 Aquisição de dados instrumentais

4.2.1 Espectroscopia NIR

As medições espectrais foram realizadas usando um Antaris MX FT-NIR espectrofotômetro (Thermo Fisher Scientific Inc, EUA) equipado com sonda de fibra óptica para transflectância. Os espectros NIR foram registrados em uma faixa de 10 000 - 4000 cm^{-1} , ou 1000-2500 nm, e foram obtidos com uma resolução espectral de 32 cm^{-1} , com 32 *scans* de varreduras. A medição do tempo foi de 26 s (32 varreduras) por espectro. 48 horas após os respectivos tratamentos de 21 e 30 dias, os espectros NIR foram adquiridos posicionando a sonda de transflectância no crânio diretamente acima do local do defeito crítico e um espectro foi adquirido para crânio, dando um total de 45 espectros

NIR. As medições espectrais foram feitas em uma sala climatizada sob temperatura controlada de 22 ° C e 60% de umidade relativa do ar.

4.2.2 Espectroscopia ATR-FTIR

As medições espectrais foram realizadas utilizando um espectrofotômetro de marca SHIMADZU, Japão, modelo IRAffinity-1, equipado com o acessório ATR. Os espectros ATR-FTIR foram obtidos numa faixa de 700 cm^{-1} a 4000 cm^{-1} , com uma resolução espectral de 4 cm^{-1} , a medição do tempo foi de 34 s (32 varreduras) por espectro. O ar foi utilizado com branco para todas as amostras. Uma gota de plasma foi adicionada diretamente sobre o cristal de ATR e um espectro foi obtido, em seguida o cristal foi limpo com álcool 70% e outra gota do mesmo plasma foi adicionada e obtida um novo espectro. Esse procedimento foi realizado em triplicata dando um total de 522 espectros ATR-FTIR sendo 276 controles e 246 sarcopênicos. As medições espectrais foram realizadas com temperatura controlada de 22 ° C e 60% de umidade relativa do ar.

4.2.3 Fluorescência Molecular

Este procedimento foi realizado para a obtenção da EEM individual da classe controle e contendo flunitrazepam. Os espectros das amostras foram adquiridos com um espectrofluorômetro RF-5301 Shimadzu, Japão usando uma cubeta de quartzo de 0,5 mm. As larguras de fenda do monocromador de excitação e emissão foram fixadas em 1,5 e 3 nm, respectivamente. As amostras foram adicionadas à cubeta com uma micropipeta de 100 μL . A cubeta foi limpa por água ultra pura depois de cada medição. A temperatura foi mantida a 25 ° C ao longo do experimento. Para amostras de extrato de larva, as superfícies espectrais de emissão / excitação foram obtidas na faixa de excitação de 280 a 390 nm (passos de 10 nm) e na faixa de emissão de 350 a 900 nm (passos de 1 nm). Este protocolo resultou em um tamanho de matriz de dados de 9 x 322 variáveis para cada amostra.

4.2.4 Voltametria de Pulso diferencial

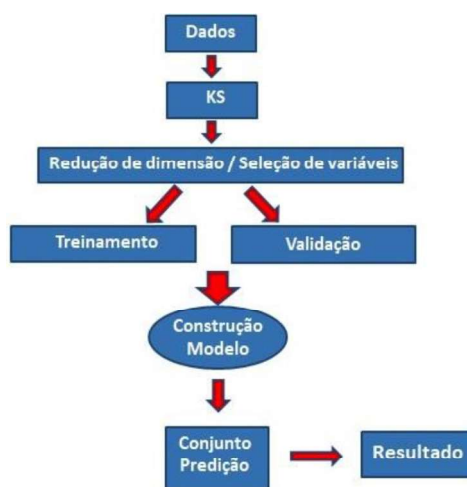
Todas as medidas de VPD foram realizadas por um bipotenciostato / galvanostato portátil μStat 400 (DropSens, Espanha), controlado pelo software DropView e realizado em tampão fosfato 0,1 mol L^{-1} a $\text{pH} = 7$, de +1,0 V para -1,0 V, taxa de varredura de 15 mV s^{-1} , tempo de pulso de 50 ms e amplitude de pulso de 100 mV. O sistema de três eletrodos foi composto por um eletrodo de referência Ag / AgCl, um fio de grafite como

eletrodo auxiliar e um eletrodo de pasta de nanotubos de carbono como eletrodo de trabalho. A composição seguinte foi utilizada na preparação do eletrodo de trabalho: 55% m:m de nanotubos de carbono de parede múltipla (Sigma-Aldrich®) e 45% m:m de óleo mineral (Specsol®). O eletrodo foi montado compactando a pasta de nanotubos de carbono na ponta de uma seringa de plástico de 1 mL. O contato elétrico interno foi feito por um fio de cobre limpo, que também atua como um êmbolo para extrusão de pasta. A renovação da superfície do eletrodo foi conseguida por extrusão de pasta e suavização em um papel de sulfito antes de cada nova medida. Todos os voltamogramas foram obtidos em triplicata.

4.3 Análise computacional

Todo o tratamento computacional, desde a importação dos dados espectrais até a construção dos modelos multivariados, foi realizado através do software MATLAB® R2014 (Mathworks, Natick, USA) usando o PLS Toolbox version 7.9.3 (Eigenvector Research, Inc., Manson, USA). Diversos pré-processamentos foram aplicados em função do tipo de dado espectral utilizado, incluindo: correção de linha de base, normalização, correção de espalhamentos de luz (Rayleigh/Raman) e espalhamento de luz multiplicativo e 1° derivada de Savitzky–Golay. Para construção dos modelos multivariados de classificação, os algoritmos PARAFAC, PCA-LDA/QDA, SPA-LDA/QDA, GALDA/QDA, 2D-LDA foram utilizados. Segue abaixo a Figura 9 apresentando o fluxograma em etapas da classificação multivariada.

Figura 9: Fluxograma representando as etapas para entrada de dados, seleção de amostras pelo algoritmo KS, redução de dimensionalidade e/ou seleção de variáveis e classificação multivariada.



Fonte: Autor

Referências

- [1] Naylor, S. What's in a Name? The Evolution of "P-Medicine". *J Precis Med.* 2015, 2, 15–29.
- [2] Zamir, E., Geiger, B. Molecular complexity and dynamics of cell-matrix adhesions. *Journal of Cell Science* 2001, 114, 3583-3590.
- [3] Naylor, S., Chen, J. Y. Nih Public Access. *National Institutes of health*, 2011, 7(3), 275–289.
- [4] <https://www.systemsbiology.org/about/what-is-systems-biology/>
- [5] Committee on mapping and sequencing the human genome. *Mapping and Sequencing the Human Genome Mapping and Sequencing the Human Genome.* [s.l: s.n.].
- [6] Pasquali, C.; Pharma, V. *From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and ...* *Nature Biotechnology*, 1996, 14.
- [7] Zhang, A. et al. Modern analytical techniques in metabolomics analysis. *Analyst*, 2012, 137, 293–300.
- [8] National human genome institute. *The Human Genome Project Completion: Frequently Asked Questions.*
- [9] Chandramouli, K., Qian, P. *Proteomics : Challenges , Techniques and Possibilities to Overcome Biological Sample Complexity.* 2009, 2009.
- [10] Legrain, P. et al. *The Human Proteome Project : Current State and Future Direction.* 2011, 1–5.
- [11] Pasquini, C. *Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications.* *J. Braz. Chem. Soc.*, 2003, 14(20), 198-219.
- [12] Neves, A. C. O., Soares, G. M., de Moraes, S. C., da Costa, F. L. S., Porto, D. L.; Lima, K. M. G. Dissolution testing of isoniazid, rifampicin, pyrazinamide and ethambutol tablets using near-infrared spectroscopy (NIRS) and multivariate calibration. *Journal of Pharmaceutical and Biomedical Analysis, Brasil*, 2012, 57, 115– 119.
- [13] Santos, M. C. D., Moraes, C. L. M., Nascimento, Y. M., Araujo, J. M. G., Lima, K. M.G. Spectroscopy with computational analysis in virological studies: A decade (2006-2016). *Trends in Analytical Chemistry* 2017, 97 244-256.
- [14] Neves, A. C. O., Moraes, C. L. M. Mendes, T. P. P., Vaz, B. G., Lima, K. M. G. Mass spectrometry and multivariate analysis to classify cervical intraepithelial neoplasia from blood plasma: an untargeted lipidomic study. *Scientific Reports* 2018, 8, 3954.

- [15] Skoog, D. A.; Holler, F. J.; Nieman, T. A.; *Princípios de Análise Instrumental*. 5^a Ed., Bookman, Porto Alegre, 2002.
- [16] Burns, D. A., Ciurczak, E. W. *Handbook of Near-Infrared Analysis*, Marcel Dekker, Inc.: New York, 1992.
- [17] Bázár, G., Kovacs, Z., Tanaka, M., Furukawa, A., Nagai, A., Osawa, M., Itakura, Y., Sugiyama, H., Tsenkova, R. Water revealed as molecular mirror when measuring low concentrations of sugar with near infrared light. *Analytica Chimica Acta*. 2015, 896, 52-62.
- [18] Mabood, F., Jabeen, F., Hussain, J., Al-Harrasi, A., Hamaed, A., Saaida, A., Al Mashaykhi, A., Al Rubaiey, Z. M. A., Manzoor, S., Ajmal Khan, Q. M., Imranul. H., Gilani, S. A., Khan, A. FT-NIRS coupled with chemometric methods as a rapid alternative tool for the detection & quantification of cow milk adulteration in camel milk samples. *Vibrational Spectroscopy*, 2017, 92, 245–250.
- [19] Materazzi, S., Peluso, G., Ripani, L., Risoluti, R. High-throughput prediction of AKB48 in emerging illicit products by NIR spectroscopy and chemometrics. *Microchemical Journal*, 2017, 134, 277–283.
- [20] Liu, T., Conde, J., Lipinski, T., Bednarkiewicz, A., Huang, C. Smart NIR linear and nonlinear optical nanomaterials for cancer theranostics: Prospects in photomedicine. *Progress in Materials Science*, 2017, 88, 89–135.
- [21] Santos, M. C., Nascimento, Y. M., Araújo, J. M. G., Lima. K. M. G. ATR-FTIR spectroscopy coupled with multivariate analysis techniques for the identification of DENV-3 in different concentrations in blood and serum : a new approach. *RSC Advances* 2017, 25640–25649.
- [22] Heraud, P.; Tobin, M. J. The emergence of biospectroscopy in stem cell research. *Stem Cell Research*, 2009, 3(1), 12–14.
- [23] Kelly, J. G. et al. Biospectroscopy to metabolically profile biomolecular structure : a multistage approach linking computational analysis with biomarkers. 2011, 1437–1448.
- [24] Potier, O. Spectral analysis and fingerprinting for biomedica characterisation. 2004, 113, 211–230.
- [25] Martin, F. L.; Obinaju, B. E.; Martin, F. L. Novel biospectroscopy sensor technologies towards environmental health monitoring in urban environments health monitoring in urban environments. *Environmental Pollution*, 2015, 183, 46–53.
- [26] Ramer, G., Lendl, B. Attenuated Total Reflection Fourier Transform Infrared Spectroscopy, 2013. doi:10.1002/9780470027318.a9287.

- [27] Skoog, D. A., Holler, F. J., Nieman, T. A. *Princípios de Análises Instrumental*, Quinta edição, Mc, Madrid, 2001. doi:10.1017/CBO9781107415324.004.
- [28] Links, D. A. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. 2012, 3202–3215.
- [29] Kuligowski, J. et al. Talanta Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics. *Talanta*, 2014, 127, 181–190.
- [30] Lima, K. M. G, Gajjar, K., Valasoulis, G., Nasioutziki, M., Kyrgiou, M., Karakitsos, P., Paraskevaidis, E., Hirschac , P. L. M., Martin, F. L.. *Analytical Methods Classification of cervical cytology for human papilloma virus (HPV) infection using biospectroscopy and variable selection*. 2014, 6, 9643–9652.
- [31] Taha, M. et al. Use of Fourier transform infrared spectroscopy (FTIR) spectroscopy for rapid and accurate identification of Yeasts isolated from human and animals. *International Journal of Veterinary Science and Medicine*, 2013, 1(1), 15–20.
- [32] Davis, R.; Mauer, L. J. Fourier transform infrared (FT-IR) spectroscopy : A rapid tool for detection and analysis of foodborne pathogenic bacteria. 2010, 1, 1582–1594.
- [33] Schmitt, J.; Flemming, H. *FTIR-spectroscopy in microbial and material analysis*. 1998, 41.
- [34] Tsutae, F. M. *Espectroscopia de correlação de fluorescência aplicado em estudos de sistemas moleculares, biológicos e celulares*. Tese (Doutorado em física) – Instituto de física de São Carlos, Universidade de São Paulo. São Paulo, p. 96. 2016.
- [35] Menezes, A. C. O. N. *Identificação e rastreamento de câncer através da combinação de análise multivariada e técnicas bioespectroscópicas*. Tese (Doutorado em Química) – Instituto de Química, Universidade Federal do Rio Grande do Norte. Rio Grande do Norte, p. 114. 2017.
- [36] Marquezin, C. A. *Técnicas de fluorescência no monitoramento de membranas modelo*. Tese (Doutorado em Física) – Faculdade de Filosofia, Ciências e Letras, Universidade de São Paulo. São Paulo, p. 97. 2008.
- [37] Trevisan, M. G. *Aplicação de Métodos Quimiométricos de Ordem Superior e Fluorescência Molecular na Análise em Matrizes Biológicas*. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Estadual de Campinas. São Paulo, p. 101. 2003.

- [38] Hess, S. T., Huang, S., Heikal, A. A., Webb, W. W. Biological and Chemical Applications of Fluorescence Correlation Spectroscopy: A Review. *Biochemistry*, 2002, 41(3), 697–705.
- [39] Fellman, J. B., Hood, E., Spencer, R. G. M. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. *Limnol. Oceanogr*, 2010, 55(6), 2452–2462.
- [40] Sádecká, J., Tóthová, J. Fluorescence spectroscopy and chemometrics in the food classification – a review. *Czech J. Food Sci.*, 2007, 25, 159–173.
- [41] Amaro, M., Brezovsky, J., Kováčová, S., Sykora, J., Bednár, D., Nemec, V., Lisková, V., Kurumbang, N. P., Beerens, K., Chaloupková, R., Paruch, K., Hof, M., Damborský, J. Site-Specific Analysis of Protein Hydration Based on Unnatural Amino Acid Fluorescence. *J. Am. Chem. Soc.* 2015, 137, 4988–4992.
- [42] Neves, A.C.O., Júnior, R. F. A., Oliveira, A. L. C. S. L., Araújo, A. A., Lima, K. M. G. The use of EEM fluorescence data and OPLS/UPLSDA algorithm to discriminate between normal and cancer cell lines: a feasibility study. *Analyst*, 2014, 139, 2423.
- [43] Bard, A. J.; Faulkner, L. R., *Electrochemical Methods: Fundamentals and Applications*. 2^a ed. New York: John Wiley & Sons, INC, 2001.
- [44] Wang, J. *Analytical Eletrochemistry*. 2^a ed. New York: John Wiley, 2000.
- [45] Souza, D. de; Machado, S. A. S.; Avaca, L. A. Voltametria de onda quadrada. Primeira parte: aspectos teóricos. *Química Nova*, 2003, 26, 81.
- [46] Brett, A. M. O.; BretT, C. M. A. *Eletroquímica princípios, métodos, e aplicações*. Coimbra: Livraria Almedina, 1996.
- [47] Zoski, C. G.; *Handbook of eletrochemistry*. 1a Ed. Netherlands: Elsevier, 2006.
- [48] Ferancova, A., Hattuniemi, M. K., Sesay, A. M., Raty, J. P., Virtanen, V. T. Electrochemical Monitoring of Nickel(II) in Mine Water. *Mine Water Environ* 2016, 35, 547–552.
- [49] Lubeckyj, R. A., Winkler-Moser, J. K., Fhaner, M. J. Application of Differential Pulse Voltammetry to Determine the Efficiency of Stripping Tocopherols from Commercial Fish Oil. *J. Am. Oil Chem. Soc.* 2017 94:527–536.
- [50] Silva, F. L. O., Tinoco, L. M. S., Ferreira, L. A. M., Malagutti, A. R., Carneiro, G. Determination of all-trans retinoic acid loaded in solid lipid nanoparticles by differential pulse voltammetry at glassy carbon electrode. *Electrochimica Acta* 2015, 182, 929–934.
- [51] Niazi A, Leardi R. Genetic algorithms in chemometrics. *J Chemom.* 2012;26(6):345-

351. doi:10.1002/cem.2426.

[52] Pontes, M. J. C. Algoritmo das Projeções Sucessivas para Seleção de Variáveis Espectrais em Problemas de Classificação. Tese (Doutorado em Química) – Departamento de Química, Universidade Federal da Paraíba. Paraíba, p. 144. 2009.

[53] Dixon, S.J.; Brereton, R.G. Comparison of performance of five common classifiers represented as boundary methods: Euclidean Distance to Centroids, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Learning Vector Quantization and Support Vector Machines, as dependent on data structure. *Chemometrics and Intelligent Laboratory Systems* 2009, 95, 1-17. DOI:10.1016/j.chemolab.2008.07.010.

[54] Bro, R. PARAFAC. Tutorial and applications, *Chemom. Intell. Lab. Syst.* (1997) 149–171. doi:10.1016/S0169-7439(97)00032-4.

[55] Sena, M.M., Trevisan, M.G., Poppi, R.J. PARAFAC: Uma ferramenta quimiométrica para tratamento de dados multidimensionais. Aplicações na determinação direta de fármacos em plasma humano por espectrofluorimetria, *Quim. Nova*. 2005, 28, 910–920. doi:10.1590/S0100-40422005000500032.

[56] Tirapelli, C. R., Fukada, S. Y., Yogi, A., Chignalia, A. Z., Tostes, R. C., Bonaventura, D., Lanchote, V. L., Cunha, F. Q., Oliveira, A. M. Gender-specific vascular effects elicited by chronic ethanol consumption in rats: a role for inducible nitric oxide synthase. *Br. J. Pharmacology* 2008, 153(3), 468-479.

[57] Shinomiya, K., Shigemoto, Y., Omichi, J., Utsu, Y., Mio, M., Kamei, C. Effects of three hypnotics on the sleep-wakefulness cycle in sleep-disturbed rats. *Psychopharmacology*, 2004, 173(1), 203-209.

Capítulo 2

Evaluation of the bony repair in rat cranial defect using near infrared reflectance spectroscopy and discriminant analysis.

Leomir A. S. de Lima

Lana S. de S. Oliveira

Aurigena A. Araújo

Raimundo Fernandes de Araújo Junior

Kássio M. G. Lima

Biotechnol. Prog., 2017, Vol. 33, No. 4, 1160 – 1167.

Contribuição:

- Realizei a aquisição espectral;
- Realizei o processamento dos dados e construção dos modelos multivariados;
- Escrevi a primeira versão do manuscrito.

Leomir A. S. de Lima

Prof. Kássio M. G. Lima

Evaluation of the Bony Repair in Rat Cranial Defect Using near Infrared Reflectance Spectroscopy and Discriminant Analysis

Leomir A. S. de Lima

Div. of Analytical Chemistry, Inst. of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio Grande do Norte, Natal, RN 59072-970, Brazil

Kássio M. G. Lima

Div. of Analytical Chemistry, Inst. of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio Grande do Norte, Natal, RN 59072-970, Brazil

Lana S. de S. Oliveira

Dept. of Biophysics and Pharmacology, Post-graduation programme in Public Health/Post graduation programme in Pharmaceutical Science, Federal University of Rio Grande do Norte, Natal, RN 59072-970, Brazil

Aurigena A. Araújo

Dept. of Biophysics and Pharmacology, Post-graduation programme in Public Health/Post graduation programme in Pharmaceutical Science, Federal University of Rio Grande do Norte, Natal, RN 59072-970, Brazil

Raimundo Fernandes de Araújo Junior

Dept. of Morphology, Post-graduation programme in Health Science/Post graduation programme in Structural and Functional Biology, Federal University of Rio Grande do Norte, Natal, RN 59072-970, Brazil

DOI 10.1002/btpr.2476

Published online April 17, 2017 in Wiley Online Library (wileyonlinelibrary.com)

We set out to determine whether near infrared reflectance spectroscopy (NIRS) combined with principal component analysis–linear discriminant analysis (LDA) or, variable selection techniques employing successive projection algorithm or genetic algorithm (GA) could evaluate the bone repair in cranial critical-size (5 mm) defect after stimulation with collagen sponge scaffold and/or infrared low-level laser therapy directly on the local. Forty-five Wistar rats were divided into nine groups of five each, namely: group H – healthy, n = 5 (without treatment and without cranial critical-size defect); (GI positive control – n = 5, 21 days or n = 5, 30 days) without treatment and with cranial critical-size defect; (GII–n = 5, 21 days or n = 5, 30 days) cranial critical-size defect filled with collagen sponge scaffold; (GIII–n = 5, 21 days or n = 5, 30 days) cranial critical-size defect submitted to low-level laser therapy; (GIV–n = 5, 21 days or n = 5, 30 days) cranial critical-size defect submitted to combined collagen sponge scaffold + low-level laser therapy treatment. In relation to the histological analysis, the collagen sponge scaffold + low-level laser therapy treatment group (GIV) 30 days showed the best result with the presence of secondary bone, immature bone (osteoid) and newly formed connective tissue (periosteum). GA–LDA model also successfully classified control class of the others classes. Thus, the results provided by the good-quality classification model revealed the feasibility of NIRS for application to evaluation of the wound healing in rat cranial defect, thanks to the short analysis time of a few seconds and nondestructive advantages of NIRS as an alternative approach for bone repair purposes. © 2017 American Institute of Chemical Engineers Biotechnol. Prog., 33:1160–1168, 2017

Keywords: cranial critical-size defect, near infrared reflectance spectroscopy, animals, classification analysis

Introduction

Rat cranial bone defects are usually with autologous grafts such as split-calvarial bone grafting,¹ iliac bones,² and costal ribs along with alloplastic.^{3–5} These materials are adequate to serve as space-maintaining materials with osteoconductive potential showing some advantages like unlimited supply, easy sterilization, and storage. Nowadays, rat bone increment

stimulus has been achieved with new applications such as chemical stimuli,⁶ biomaterials,⁷ collagen sponge scaffold,⁸ and the use of physical stimuli such as ultrasound,⁹ electromagnetic fields,¹⁰ and low-level laser therapy.¹¹ These stimuli are applied for osteoconductive, osteoinductive, or osteogenetic purposes.

There are a several methods used for evaluation of the growth of rat bone in bony defects is histological analysis such as histomorphometry (HM),¹² radiography,¹³ and micro-computed tomography (μ CT).¹⁴ Although histological evaluation of bone growth inside the defect is considered the

Correspondence concerning this article should be addressed to K. M. G. Lima at kassiolima@gmail.com.

gold standard, it has drawbacks to this technique. One of which is that the histological section shows only one specific are of the defect or cross-section, which does not necessarily represent the entire defect. Furthermore, traditional techniques such as HM is time consuming and rather expensive.

Alternatively, the near-infrared reflectance spectroscopy (NIRS) and others techniques (Raman, IR) are among the noninvasive methods, emerging as an interesting alternative for measure a variety of biomedical applications in rats^{15–20} using the wavelength range of 900 to 2,500 nm. NIRS is also characterized by a minimum of sample handling. It requires no extractions and is nondestructive. The NIR absorption is affected by the internal and external biochemical composition of the organism.

For example, Radhakrishnan et al.¹⁵ quantified reduced light scattering coefficient of near-IR light in the rat spinal cord region and to differentiated healthy and demyelinated peripheral nerves intraoperatively based on differential light scattering. NIR radiation offers the advantage of being nonionizing, and therefore, repeated doses can be given to the subject. Olesberg et al.¹⁶ performed in vivo measurements of near-infrared rat skin absorption during a glucose clamp experiment, identifying the presence of glucose-specific spectral information. Hull et al.¹⁷ evaluated the ability of steady-state, radially resolved, broad-band near infrared diffuse reflectance spectroscopy to measure carbogen-induced changes in haemoglobin oxygen saturation (SO₂) and total haemoglobin concentration in a rat R3230 mammary adenocarcinoma model in vivo. The authors concluded that near infrared diffuse reflectance spectroscopy provides a rapid, noninvasive means of monitoring spatially averaged changes in tumor haemoglobin oxygen saturation induced by oxygen modifiers. Lastly, Neves et al.¹⁸ developed a quantitative method for analysis of glucose, triglycerides, and HDL in rat plasma without sample pretreatment, using direct NIR absorption measurements. An important aspect of this approach is that no reagents are necessary except for the aqueous solutions used to clean and rinse the optical cell. In addition to increasing the speed of the analysis, the proposed method may also provide a reduction in waste production compared to traditional methods. Goldfeld et al.¹⁹ employed the NIRS combined with multivariate calibration models to monitor the concentrations of glycerol and methanol in real time during the production of monoclonal antibody by a *Pichia pastoris* high cell density process. Finally, Berry²⁰ et al. described the multiscale Raman spectroscopy modeling from a monoclonal antibody producing CHO cell culture process including data from two development scales (3 L, 200 L) and a clinical manufacturing scale environment (2,000 L).

Conversely, NIRS generates large datasets with hundreds of variables. Multivariate (calibration and classification) analysis needs to be employed to determine the significant but subtle differences in the spectra. Nondestructive classification is one of the interesting applications for biomedical applications. Many studies have used principal component analysis (PCA) for initial data reduction,²¹ hierarchical cluster analysis (HCA) for analyse groups in a set of data on the basis of spectral similarities,²² and linear discriminant analysis (LDA) for classify unknown samples into predetermined groups.²³ LDA as a classifier, with variable selection by the successive projections algorithm (SPA)²⁴ and genetic algorithm (GA)²⁵ have been successfully used for this purpose in different classification problems.

Although these studies have shown that NIRS and discriminant analysis are an interesting alternative for measure a variety of biomedical applications in rats, little research

has been directed toward the use of NIRS to evaluate the bone repair in defects created in rat cranium. Herein, we have attempted to evaluate the potential of a novel nondestructive method for evaluation of the wound healing in rat cranial defect as an alternative approach for bone repair purposes. For this, the present article proposes the NIRS region, or combination of variables, that reflects a specific biochemical feature of bone increment stimulus implanted beneath the cranial periosteum of rats to observe bone development and absorbent change of healthy cranial, cranial critical-size defect, cranial critical-size defect treatment with collagen sponge, cranial critical-size defect treatment with low-level laser therapy and cranial critical-size defect treatment with collagen sponge + low-level laser therapy. We employed SPA and GA to select an appropriate subset of wavenumbers for LDA. Other goals were the elucidation of the altered variables using different bone increment stimulus. This novel approach can lead to more selective and specific identification as a new method of inducing bone formation or influence bone graft modeling. Nevertheless, bone increment stimulus was never calibrated by NIRS using discriminant analysis.

Experimental

The study protocol was approved by the Ethics in Animal Research Committee of Federal University of Rio Grande do Norte, Brazil (Process number 067/2014) and the study design was approved. Male adult Wistar rats were obtained from the vivarium of the Department of Biophysics and Pharmacology of UFRN, Brazil. All animals were housed in an animal room under standard laboratory conditions at $22 \pm 2^\circ\text{C}$ with a 12-h/12-h light/dark cycle. Animals were fed pelleted food and water *ad libitum*. They were acclimatized for 7 days and fasted for 12 h before the experiments. All efforts were made to minimize the number of animals used and their suffering degree.

Animal model

Forty-five male, adult Wistar rats weighting 320 g on average were used in this study. They were divided into nine groups of five each, namely: group H – healthy, $n = 5$ (without treatment and without cranial critical-size defect), (GI – $n = 5$, 21 days or $n = 5$, 30 days) without treatment and with cranial critical-size defect (positive control; GII – $n = 5$, 21 days or $n = 5$, 30 days) cranial critical-size defect filled with collagen sponge scaffold; (GIII – $n = 5$, 21 days or $n = 5$, 30 days) cranial critical-size defect submitted to low-level laser therapy (wavelength 780 nm; total energy density 120 J cm^{-2} ; power 50 mW); (GIV – $n = 5$, 21 days or $n = 5$, 30 days) cranial critical-size defect submitted to combined collagen sponge scaffold + low-level laser therapy treatment. Each group will create a cranial critical-size defect (5 mm), except H – healthy group. The groups received the incidence laser diode (Kondor-tech - Biowave LLLT Dual, Brazil) at a dose of 120 J cm^{-2} (power 50 mW, exposure time 10 s) and wavelength of 780 nm, with applications made in 4 points equally spaced around the faulty distance of 1 mm from the edge, soon after surgery. The collagen sponge used was the brand Hemospon® (Tecnew 10 Cat.). It was cut with a scalpel blade at the approximate size of the defect, with a thickness of approximately 1mm.

Surgical procedure

The animals were anesthetized with an intraperitoneal injection of ketamine hydrochloride (10% v/v) at a dose of

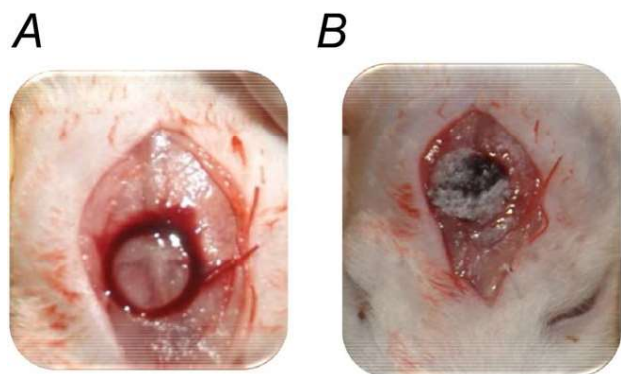


Figure 1. (A) cranial critical-size (5 mm) defect. (B) cranial critical-size (5 mm) defect with a collagen sponge scaffold.

Table 1. Ray Score X

Scores	Definition
0	The defect has 0–25% of its total area with image radiopaque visible, consistent with mineralized bone tissue
1	The defect has 25.1 to 50% of its total area with image radiopaque visible, consistent with mineralized bone tissue
2	The defect has 50.1 to 75% of its total area with image radiopaque visible, consistent with mineralized bone tissue;
3	The defect has 75.1 to 100% of its total area with visible radiopaque image, compatible with bone tissue mineralized.

80 mg kg⁻¹ and Xylazine (2% v/v) at a dose of 10 mg Kg⁻¹. A skin incision made from the front nose area in the external occipital bone lump, exposing the extent of the calvarial (Figure 1A). A cranial critical-size defect of 5 mm was made between the two parietal regions in the cranium center with a trephine drill driven by an electric micromotor 30,000 rpm (SMART, Driller, São Paulo, Brazil) under constant irrigation of sterile saline solution to avoid overheating bone.²⁶ In the groups GII (collagen sponge scaffold treatment) and GIV (collagen sponge scaffold + low-level laser therapy treatment), the cranial critical-size defect were immediately filled with a collagen sponge (Hemospon® – Technew, São Paulo, Brazil; Figure 1B). In the groups GIII (low-level laser therapy) and GIV (collagen sponge scaffold + low-level laser therapy treatment), the cranial critical-size defect areas were irradiated with a GaAlAs diode laser (MMOptics, Brazil) in continuous mode using a power of 50mW, wavelength of 780 nm, and total energy density of 120 J cm⁻². Applications were made at four equidistant points 1 mm distant from the edge of the defect, immediately after surgery and at intervals of 48 and 96 h. For postoperative care was used analgesic dipyrone monohydrate 500 mg, a drop oral for 5 days, and 5 drops into the trough for 1 week.

Radiography

After the experimental intervals the animals were submitted to euthanasia (thiopental, 2%, 100 mg/kg, Thiopentax Cristália, São Paulo, Brazil). Radiographic analyzes were performed at the Federal University of Rio Grande do Norte, in the Odontology department using Ray-x Timex 70c, ice wall 220 v + 4% –60Hz, Gnauts, Brazil). The animal cranium of

Table 2. Number of Training, Validation, and Prediction Specimens in Each Category

Category	Set Training	Set Validation	Set Prediction
Healthy (H)	3	1	1
Positive Control (CI)	6	2	2
Collagen sponge scaffold (CII)	6	2	2
Low-level laser therapy (CIII)	6	2	2
Collagen sponge scaffold + low-level laser therapy (CIV)	6	2	2

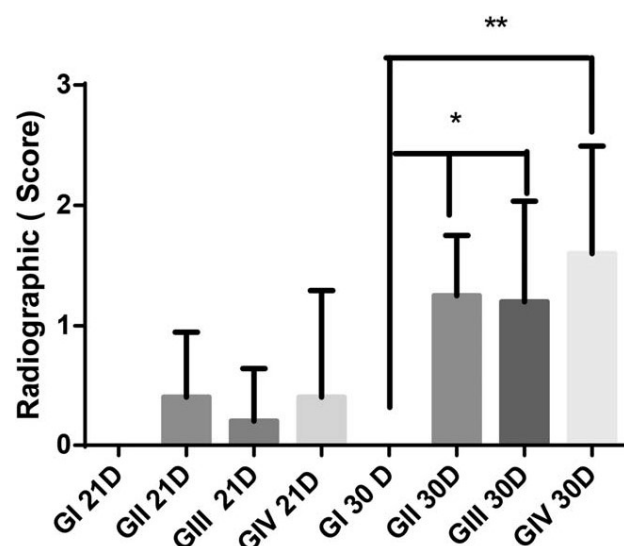


Figure 2. Bone repair of groups using radiographic analysis.

No significant difference among the groups 21 days. Significantly difference among treated groups 30 days (GII and GIII, * $P < 0.05$; GIV, ** $P < 0.01$) when compared to Positive control (GI 30 days). GI = 21D Positive Control 21 days; 21D GII = 21 days of collagen sponge scaffold; 21D GIII = 21 days low-level laser therapy; GIV 21D = collagen sponge scaffold + low-level laser therapy 21 days therapy; GI = 30D Positive Control 30 days; GII 30D = collagen sponge scaffold to 30 days; GIII 30D low-level laser therapy 30 days; GIV 30D collagen sponge scaffold + low-level laser therapy 30 days Kruskal–Wallis test followed by Dunn's test.

copies and the x-ray exposure time was 0.32, the pieces were 40 cm of equipment were taken. For development, we used 90 seconds in the developer, 30 seconds in water and 3 minutes in fixative. Radiographs were scanned, suffered no process of image modification, were evaluated using an imaging program, Image Tolls, and given scores through the analysis of an examiner, based on the table below (Table 1).

Histological assessments

For histology, cranial specimens were fixed in 10% formaldehyde decalcifying with 5% (v/v) nitric acid was made for 10 days. Then it was dehydrated with three alcohol baths (80%, 90%, and 100%) for 1 h each, followed by three baths of xylene 1 h each, and finally three baths paraffin, followed by “embedment” parts and identification. After the blocks were correctly identified 5 μ m cuts made on the micrometer and these fixed sections on slides and stained with hematoxylin and eosin.²⁷ Were evaluated by an experienced examiner, the following histopathologic events: (A) the degree of inflammation; (B) Training and quality of bone tissue formed; (C) and degree of maturation of collagen.

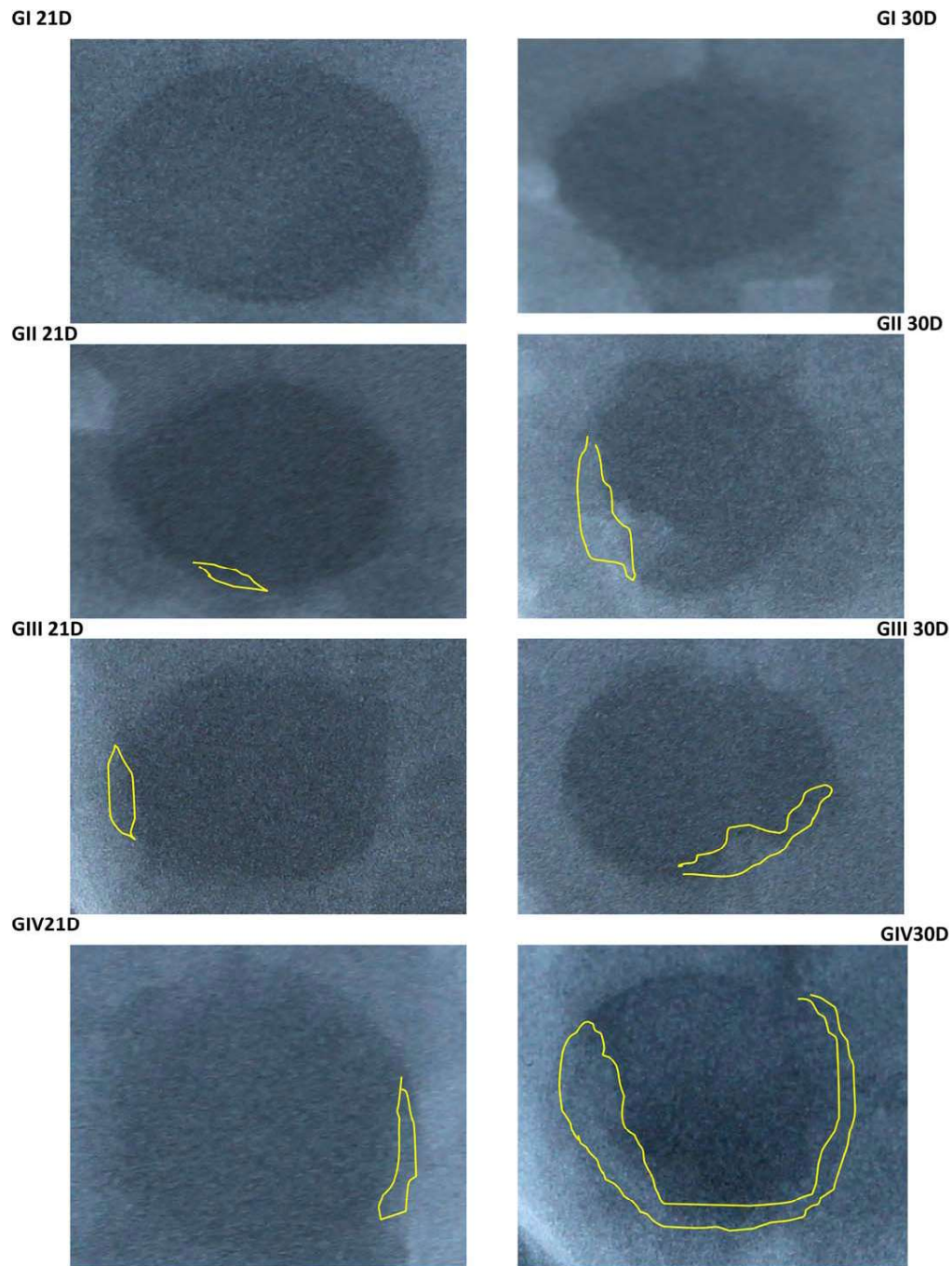


Figure 3. Radiographic images of each group.

Yellow line (radiopaque areas, neo formation bone). Positive control (GI). GI = 21D Positive Control 21 days; 21D GII = 21 days of collagen sponge scaffold; 21D GIII = 21 days low-level laser therapy; GIV 21D = collagen sponge scaffold + low-level laser therapy 21 days therapy; GI = 30D Positive Control 30 days; GII 30D = collagen sponge scaffold to 30 days; GIII 30D low-level laser therapy to 30 days; GIV 30D collagen sponge scaffold + low-level laser therapy 30 days.

The stained sections were photographed digitally under a microscope (Digital Camera DXM200F, Nikon, Japan). The image analysis was used to estimate the new forming bone at 21 days and 30 days to calculate the percentage of newly formed bone within the defect using the computerized image analysis software Image-J 6.0. Briefly, the image was corrected for the optical density before the region-of-interest (ROI) in mm was selected in accordance with the color specific for the region of new bone

formation in the histological slide. Then the parameters of the ROI, such as area, were calculated. Automation of these steps is included in an algorithm referred as a macro. The macro was used to normalize the selection of other ROIs. The percentage of newly formed bone was determined by ROI area/total area of images. Three sections from each sample were used for semi-quantitative image analysis. Three regions in each section were photographed at 200x magnification

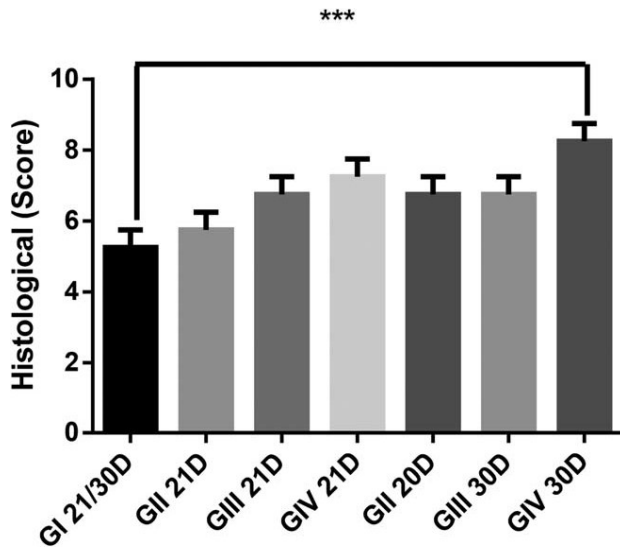


Figure 4. Bone repair of the groups 21 days and 30 days from histological analysis.

There was a significant difference between the positive control group (GI) and GIV 30D collagen sponge scaffold + low-level laser therapy 30 days, *** $P < 0.001$. Kruskal–Wallis test followed by Dunn's test.

Statistical analysis

Data are presented as the mean \pm standard error of the mean or as the median (range), when appropriate. Analysis of variance (ANOVA) followed by Bonferroni's test was used to calculate the means. The Kruskal–Wallis test followed by Dunn's test was used to compare medians (Graph Pad Prism 5.0 Software, La Jolla, CA). A P -value < 0.05 indicated a statistically significant difference.

NIR spectra measurement

Spectral measurements were performed using an Antaris MX FT-NIR spectrophotometer (Thermo Fisher Scientific Inc) equipped with transreflectance fiber optic probe. The NIR spectra were obtained over a range of 10,000–4,000 cm^{-1} , or 1,000–2,500 nm, and were recorded with a spectral resolution 32 cm^{-1} , with 32 scans co-added. The time measurement was of 26 s (32 scans) per spectrum. 48 h after the respective treatments of 21 and 30 days NIR spectra were acquired positioning the probe transreflectance on the cranium directly above the critical defect site and one spectra was acquired for cranium, giving in a total of 45 NIR spectra. Absorbance spectra of cranium samples were obtained against the spectrum of the spectralon used as background. Spectral measurements were done in an acclimatized room under controlled temperature of 22°C, and 60% relative air humidity.

Chemometrics procedure and software

The data import, pre-treatment and construction of chemometric classification models (PCA-LDA, SPA-LDA, and GA-LDA) were implemented in MATLAB R2014a software (Math works Inc, Natick, MA). The raw spectra were pre-processed cutting 1,250–2,500 nm, Savitzky-Golay filter (first order, 7 points) and multiplicative scatter correction (MSC) were applied. Mean centering was applied to all spectra before performing variable subset selection and calibration. For

PCA-LDA, SPA-LDA, and GA-LDA model, the samples were divided into training (70%), validation (15%), and prediction sets (15%) by applying the classic Kennard–Stone (KS) uniform sampling algorithm to the NIR spectra,²⁸ sample numbers in each set are presented in Table 2. Training samples were used in the modeling procedure (including variable selection for LDA), whereas the prediction set was only used in the final evaluation of the classification. The optimum number of variables for SPA-LDA and GA-LDA was determined from the minimum cost function G calculated for a given validation dataset:

$$G = \frac{1}{N_V} \sum_{n=1}^{N_V} g_n, \quad (1)$$

where g_n is defined as

$$g_n = \frac{r^2(x_n, m_{I(n)})}{\min_{I(m) \neq I(n)} r^2(x_n, m_{I(m)})} \quad (2)$$

and $I(n)$ is the index of the true class for the n th validation object x_n . g_n is defined as risk of misclassification of the n th validation object x_n , $n=1, \dots, N_V$. In this definition, the numerator is the squared Mahalanobis distance between object x_n (of class index I_n) and the sample mean $m_{I(n)}$ of its true class. The denominator in Eq. (2) corresponds to the squared Mahalanobis distance between object x_n and the center of the closest wrong class.

GA-LDA was used to select variables employing the G function as cost function. The mutation and reproduction probabilities were kept constant, 60% and 10%, respectively. The initial population was carried out for 40 generations with 80 chromosomes each.

Results and Discussion

All the animals survived the operation and each of the postoperative examinations, and could be evaluated at the end of the experiment. The results showed no significant differences in scores for X-rays analysis in the groups treatment for 21 days compared com Positive control (GI). However, treated groups (GII, GIII, and GIV) 30 days were significantly different when compared to the Positive control group (GI) of 30 days. As the GIV group most statistical highlight with $P < 0.01$ (Figure 2).

As for the groups 30 days, GIV showed a mean 55.16% bone repair, while the GII and GIII, respectively, showed 39.14% and 35.32%. In the radiographic evaluation groups GII, GIII, and GIV 30 days were favorable when compared to the Positive control group (GI). That is, the collagen sponge scaffold + low-level laser therapy treatment; GIV) group 30 days showed significantly higher radiopaque areas than the other groups, indicating a possible formation of a mineralized matrix and possible bone formation (Figure 3). So that radiographically, the combination of the two techniques with 30 days of repair would be the best treatment for regeneration of the critical size defect in the skull of rats. At 21 days, radiographically, there was no significant difference between the treated groups and the control group, and not between the treated groups when compared to each other. GII had about 19.25% of bone repair, GIII and GIV were 12.79% and 27.91%, respectively. Thus, the GIV still 21 days presents with good results, is a good solution to a potentially faster recovery, but no significantly difference.

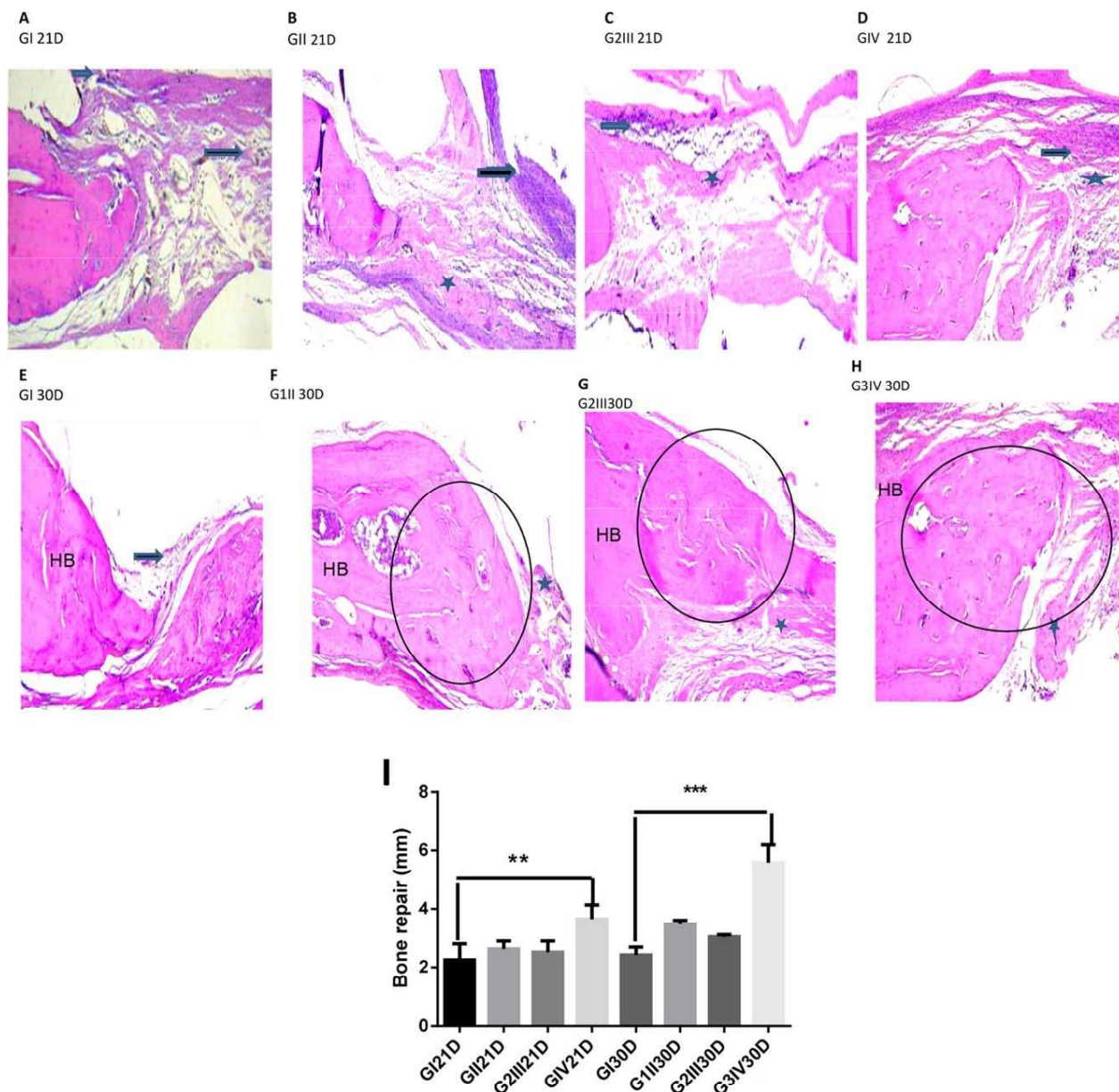


Figure 5. Histological analysis of graft areas after treatment at 21 and 30 days.

Defect areas were treated with collagen sponge scaffold (B,F), low-level laser therapy (C,G), and collagen sponge scaffold + low-level laser therapy (D,H). General images (A–D) showing inflammatory infiltrate and intense collagen formation at 21 days after treatment. Images (E–H) showing collagen produced in the newly formed tissues at 30 days after treatment. (I) Quantification of bone regeneration via morphometric analysis confirms semiquantitative study (analysis of variance followed by Bonferroni's correction). The black arrow: inflammatory infiltrate. The asterisk: collagen formation. The white circle: new bone. HB: host bone.

The results of the histological slides were significantly for collagen sponge scaffold + low-level laser therapy treatment group (GIV) 30 days when compared to the control group (GI) to 30 days, $P < 0.001$ (Figure 4). As shown in Figure 5H,I for collagen sponge scaffold + low-level laser therapy treatment group (GIV) 30 days, we can see the presence of secondary bone, immature bone (osteoid), and newly formed connective tissue (periosteum) in nearly . The low-level laser therapy groups (GIII) 21 days and collagen sponge scaffold + low-level laser therapy treatment) (GIV) 21 days not show the formation of immature bone.

At 21 days, the positive control group (GI) showed the intense presence of inflammatory cells, new tissue formation through the defect filling with connective tissue containing blood

capillaries, fibroblasts, macrophages and newly formed collagen fibers, and no evidence of union bone. After 30 days, the same group showed almost the same characteristics present on day 21, decreasing the presence of inflammatory cells and showing a slightly more organized tissue, a bone matrix (Figure 5).

Figure 6A shows the collected raw diffuse reflectance spectra of nine groups of rat cranial samples namely: group H-healthy (blue color), group CI-control positive [with cranial critical-size defect, (red color)], group CII-collagen sponge scaffold (black color), group CIII-low-level laser therapy (green color), and group CIV collagen sponge scaffold + low-level laser therapy – (blue magenta color). Baseline offsets and bias were present due to the light scattering. As can be seen in Figure 6A, given the vast quantity of data and the overlap, it is

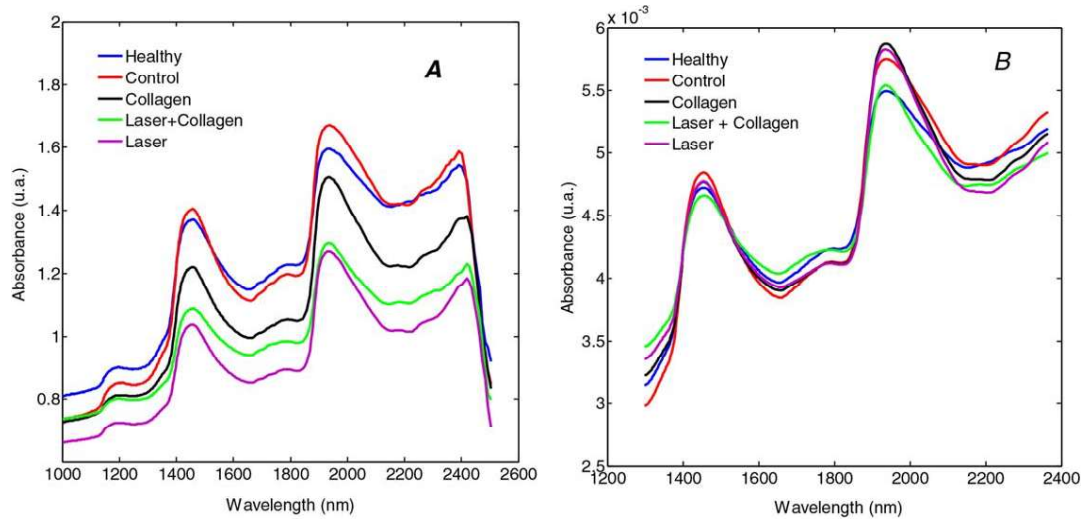


Figure 6. (A) Average spectra acquired from all categories.

The spectra from healthy are shown in blue; those with control are shown in red; those with collagen are shown in black; those with laser and collagen are shown in green and, those from laser are pink. (B) Preprocessing spectra from all categories. The spectra from healthy are shown in blue; those with control are shown in red; those with collagen are shown in black; those with laser and collagen are shown in green and, those from laser are pink.

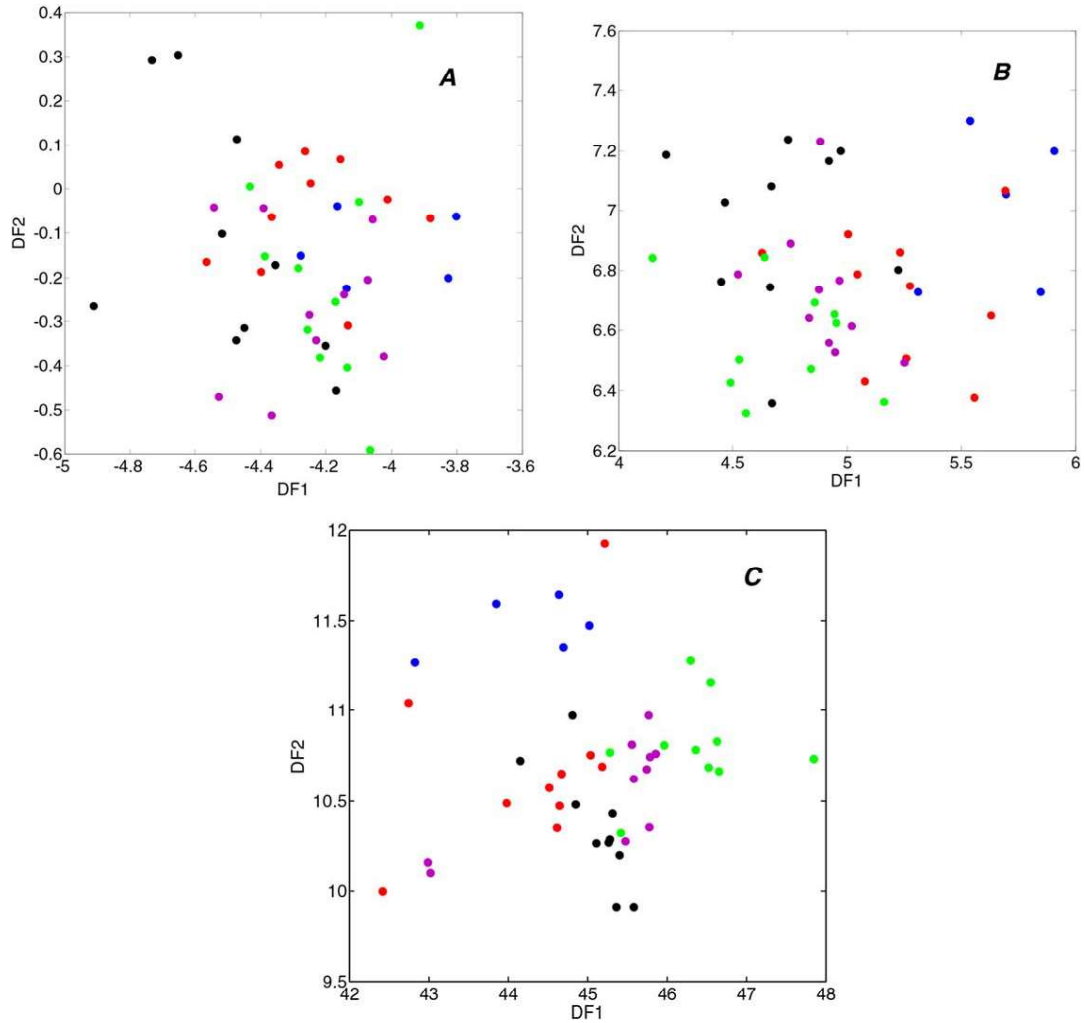


Figure 7. The application of linear discriminant analysis to the segregation of cranial bone defects categorized healthy (blue), positive control (red), collagen sponge scaffold (black), collagen sponge scaffold + low-level laser therapy (green), and low-level laser therapy (blue magenta) stimulus.

(A) DF1 × DF2 discriminant function values calculated by using principal component analysis (PCA)-linear discriminant analysis (LDA) results from all categories. (B) DF1 × DF2 discriminant function values calculated using the variables selected by successive projection algorithm (SPA)-linear discriminant analysis (LDA) results from all categories. (C) DF1 × DF2 discriminant function values calculated by using the variables selected by genetic algorithm (GA)-linear discriminant analysis (LDA) results from all categories.

difficult to distinguish inter-category (i.e., healthy vs. positive control vs. collagen sponge vs. low-level laser therapy vs. collagen sponge scaffold + low-level laser therapy) variance.

To solve undesirable features in the NIR spectra, some pretreatments were applied, such as Savitzky–Golay filter (first order, 7 points) and normalizing. Further, the raw spectra was cut in 1,250–2,400 nm, giving models with better predictive abilities than those built using whole NIR region. These effects are shown in Figure 6B.

To this end, multivariate analysis (PCA, SPA, and GA) were applied to identify wavelengths responsible for intercategory variance and this was then combined with subsequent LDA to maximize separation in scores plots while minimizing intracategory influences and also to evaluate it as an exploratory tool to increase the understanding about the differences between classes.

The best PCA-LDA result was achieved by using four PCs, accounting for 99.0% of the variance, reaching most diagnostically significant ($P < 0.05$) for discriminating each bone repair defects in rat cranium after stimulation. The Figure 7A shows the plot scores ($DF1 \times DF2$) of five categories (cranium defect), showing that there is overlapping among all categories with a minimal discrimination. For, SPA-LDA using 10 selected wavelengths (namely, 1,299 nm, 1,399 nm, 1,443 nm, 1,719 nm, 1,820 nm, 1,935 nm, 2,006 nm, 2,071 nm, 2,227 nm, and 2,365 nm), obtained by cost function G, achieved an improved segregation ($P < 0.05$) between classes (Figure 7B) when compared with PCA-LDA. However, while the healthy vs. laser showed some degree of separation, although complete overlap of the two clusters (laser vs. laser and collagen) is observed in the DF scores plot (Figure 7B). However, when GA-LDA was employed to predict all categories ($P < 0.05$), using only 17 selected wavelengths (namely, 1,347 nm, 1,361 nm, 1,388 nm, 1,403 nm, 1,427 nm, 1,514 nm, 1,588 nm, 1,692 nm, 1,710 nm, 1,725 nm, 1,964 nm, 1,988 nm, 2,019 nm, 2,071 nm, 2,125 nm, 2,160 nm, and 2,348 nm) give better segregation than PCA-LDA and SPA-LDA together, as shown in Figure 7C. In addition, it is clear that the cluster scores for GA-LDA in comparison to SPA-LDA and GA-LDA are distinct, demonstrating different wavelengths are responsible for evaluation of the wound healing in rat cranial defect.

Examination of the selected wavelengths following GA-LDA showed that the main biochemical alterations discriminating the different stimulus. Several selected wavenumbers appear to be of particular interest, namely, the variables at 1,725 nm (the first overtone of *n*-alkane of lipid band) and a protein band around 2,071 nm (*a* combination band of amide A and amide II of polypeptides) for a quantitative evaluation of collagen changes in the cranium defects, respectively.

Currently, histopathological bone tissue analysis in repair by a pathologist represents the only definitive method for confirmation of presence or absence of bone formation, and bone formation grading, or the measurement of inflammatory process. In this context, the need for analysis through spectroscopy (NIRS) combined with principal component analysis–linear discriminant analysis (PCA-LDA) or, variable selection techniques employing successive projection algorithm (SPA) or genetic algorithm (GA) have been tested to get conclusive results in the bone repair. Our results showed that near infrared reflectance spectroscopy and discriminant analysis are useful to evaluate bony tissue reconstruction due to the fact that the results provided by the good-quality classification model corroborated to histopathological findings.

In treated groups, there were collagen formation and decreased inflammation compared to nontreated group reflecting in biochemical alterations demonstrated by wavenumbers. These alterations seem be most evident in group 30 days.

Conclusions

In conclusion, the joint use of NIRS and discrimination analysis is a powerful means for evaluation of the wound healing in rat cranial defect in different stimulus. This study proposed to evaluate the potential of a novel nondestructive method for identification of five categories (healthy, positive control, collagen sponge scaffold, low-level laser therapy and collagen sponge scaffold + low-level laser therapy) as an alternative approach for cranial bone defects, employing a simple NIR spectrometry and LDA analysis coupled with the GA for wavenumber selection. Variable selection techniques (SPA and GA) using LDA was subsequently performed in an attempt to gain more information regarding potential differences among the categories from the NIR spectra. Thus, our NIR spectroscopic method would be able not only to discriminate chemical or physical stimuli in cranial bone defects but also to distinguish others applications in rats such as evaluation of traumatic brain injury.²⁹ The present study indicates that NIR spectroscopy has good performance in discriminating different stimulus in rats and it may offer potential for noninvasive, in vivo diagnosis of cranial bone defect in the near future.

Acknowledgments

Leomir A.S. de Lima would like to thank PPGQ-UFRN for scientific support. K.M.G. Lima acknowledges the CNPq/Capes project (Grant 070/2012 and 305962/2014-0), PPP 004/2012) for financial support.

Literature Cited

- Thaller SR, Hoyt J, Dart A, Borjeson K, Tesluk H. Repair of experimental calvarial defects with bio-ss particles and collagen sponges in a rabbit model. *J Craniofac Surg.* 1994;5:79–84.
- Sullivan WG, Szwajkun PR. Revascularization of cranial versus iliac crest bone grafts in the rat. *Plast Reconstr Surg.* 1991;87:1105–1109.
- Yazici M, Emans J. Fusionless instrumentation systems for congenital scoliosis:expandable spinal rods and vertical expandable prosthetic titanium rib in the management of congenital spine deformities in the growing child. *Spine (Phila. Pa. 1976).* 2009;34:1800–1807.
- Roy ME, Nishimoto SK, Rho JY, Bhattacharya SK, Lin JS, Pharr GM. Correlations between osteocalcin content, degree of mineralization, and mechanical properties of C. carpio rib bone. *J Biomed Mater Res.* 2001;54:547–553.
- Holton LH III, Chung T, Silverman RP, Haerian H, Goldberg NH, Burrows WM, Gobin A, Butler CE. Comparison of acellular dermal matrix and synthetic mesh for lateral chest wall reconstruction in a rabbit model. *Plast Reconstr Surg.* 2007;119:1238–1246.
- Ustinova EE, Fraser MO, Pezzone MA. Colonic irritation in the rat sensitizes urinary bladder afferents to mechanical and chemical stimuli: an afferent origin of pelvic organ cross-sensitization. *Am J Physiol Renal Physiol.* 2006;290:1478–1487.
- Dausse Y, Grossin L, Miralles G, Palletier S, Mainard D, Hubert P, Baptiste D, Gillet P, Dellacherie E, Netter P, Payan E. Cartilage repair using new polysaccharidic biomaterials: macroscopic, histological and biochemical approaches in a rat model of cartilage defect. *Osteoarthritis Cartil.* 2003;11:16–28.

8. de Brito Bezerra B, Mendes Brazão MA, de Campos MLG, Casati MZ, Sallum EA, Sallum AW. Association of hyaluronic acid with a collagen scaffold may improve bone healing in critical-size bone defects. *Clin Oral Implants Res.* 2012;23:938–942.
9. Angle SR, Sena K, Sumner DR, Virdi AS. Osteogenic differentiation of rat bone marrow stromal cells by various intensities of low-intensity pulsed ultrasound. *Ultrasonics.* 2011;51:281–288.
10. Chang K, Chang WHS. Pulsed electromagnetic fields prevent osteoporosis in an ovariectomized female rat model: a prostaglandin E2-associated process. *Bioelectromagnetics.* 2003;24:189–198.
11. Matsumoto MA, Ferino RV, Monteleone GF, Ribeiro DA. Low-level laser therapy modulates cyclo-oxygenase-2 expression during bone repair in rats. *Lasers Med Sci.* 2009;24:195–201.
12. Pataki A, Müller K, Green JR, Ma YF, Li QN, Jee WSS. Effects of short-term treatment with the bisphosphonates zoledronate and pamidronate on rat bone: a comparative histomorphometric study on the cancellous bone formed before, during, and after treatment. *Anat Rec.* 1997;249:458–468.
13. Boivin G, Meunier PJ. The degree of mineralization of bone tissue measured by computerized quantitative contact microradiography. *Calcif Tissue Int.* 2002;70:503–511.
14. Boyd SK, Davison P, Muller R, Gasser JA. Monitoring individual morphological changes over time in ovariectomized rats by in vivo micro-computed tomography. *Bone.* 2006;39:854–862.
15. Radhakrishnan H, Senapati A, Kashyap D, Peng YB, Liu H. Light scattering from rat nervous system measured intraoperatively by near-infrared reflectance spectroscopy. *J Biomed Opt.* 2005;10:051405. (October):
16. Olesberg JT, Liu L, Van Zee V, Arnold MA. In vivo near-infrared spectroscopy of rat skin tissue with varying blood glucose levels. *Anal Chem.* 2006;78:215–223.
17. Hull EL, Conover DL, Foster TH. Carbogen-induced changes in rat mammary tumour oxygenation reported by near infrared spectroscopy. *Br J Cancer.* 1999;79:1709–1716.
18. Neves ACDO, de Araújo AA, Silva BL, Valderrama P, Março PH, de Lima KMG. Near infrared spectroscopy and multivariate calibration for simultaneous determination of glucose, triglycerides and high-density lipoprotein in animal plasma. *J Pharm Biomed Anal.* 2012;66:252–257.
19. Goldfeld M, Christensen J, Pollard D, Gibson ER, Olesberg JT, Koerperick EJ, Lanz K, Small GW, Arnold MA, Evans CE. Advanced near-infrared monitor for stable real-time measurement and control of *Pichia pastoris* bioprocesses. *Biotechnol Prog.* 2014;30:749–759.
20. Berry B, Moretto J, Matthews T, Smelko J, Wiltberger K. Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis. *Biotechnol Prog.* 2015;31:566–577.
21. Marques AS, Nicácio JTN, Cidral TA, de Melo MCN, de Lima KMG. The use of near infrared spectroscopy and multivariate techniques to differentiate *Escherichia coli* and *Salmonella enteritidis* inoculated into pulp juice. *J Microbiol Methods.* 2013;93:90–94.
22. Martin FL, Llabjani V, Evans G, Trevisan J, Martin-Hirsch PL, Patel II, Stringfellow HF. High contrast images of uterine tissue derived using Raman microspectroscopy with the empty modelling approach of multivariate curve resolution-alternating least squares. *Analyst.* 2011;136:4950–4959.
23. Cheung KT, Trevisan J, Kelly JG, Ashton KM, Stringfellow HF, Taylor SE, Singh MN, Martin-Hirsch PL, Martin FL. Fourier-transform infrared spectroscopy discriminates a spectral signature of endometriosis independent of inter-individual variation. *Analyst.* 2011;136:2047–2055.
24. Marques AS, de Melo MCN, Cidral TA, de Lima KMG. Feature selection strategies for identification of *Staphylococcus aureus* recovered in blood cultures using FT-IR spectroscopy successive projections algorithm for variable selection: a case study. *J Microbiol Methods.* 2013;98:26–30.
25. Marques AS, Moraes EP, Júnior MAA, Moura AD, Neto VFA, Neto RM, Lima KMG. Rapid discrimination of *klebsiella pneumoniae* carbapenemase 2—producing and non-producing *klebsiella pneumoniae* strains using near-infrared spectroscopy (NIRS) and multivariate analysis. *Talanta.* 2015;134:126–131.
26. Soares LGP, Marques AMC, Guarda MG, Aciole JMS, Dos Santos JN, Pinheiro ALB. Influence of the 780 nm laser light on the repair of surgical bone defects grafted or not with biphasic synthetic micro-granular hydroxylapatite + Beta-Calcium triphosphate. *J Photochem Photobiol B Biol* 2014;131:16–23.
27. Riccio M, Resca E, Maraldi T, Pisciotto A, Ferrari A, Bruzzesi G, D, Pol A. Human dental pulp stem cells produce mineralized matrix in 2D and 3D cultures. *Eur J Histochem.* 2010;54:205–213.
28. Kennard R, Stone L. Computer aided design of experiments. *Technometric.* 1969;11:137–148.
29. Lin B-S, Wang C-C, Chang M-H, Chio C-C. Evaluation of traumatic brain injury by optical technique. *BMC Neurol.* 2015;15:202.

Manuscript received Mar. 14, 2016, and revision received Dec. 30, 2016.

Capítulo 3

Ferramenta de triagem para sarcopenia em idosos: uma abordagem bioespectroscópica.

Leomir A. S. de Lima

Daniele Sirineu Pereira

Juliana Fernandes de Souza Barbosa

Ricardo Oliveira Guerra

Kássio M. G. de Lima

Manuscrito em fase de escrita.

Contribuição:

- Realizei a aquisição espectral;
- Realizei o processamento dos dados e construção dos modelos multivariados;
- Escrevi a primeira versão do manuscrito.

Leomir A. S. de Lima

Prof. Kássio M. G. Lima

Ferramenta de triagem para sarcopenia em idosos: Uma abordagem bioespectrocópica.

Leomir Aires Silva de Lima¹, Daniele Sirineu Pereira², Juliana Fernandes de Souza Barbosa³, Aline do Nascimento Falcão Freire³, Ricardo Oliveira Guerra³, Kássio Michell Gomes de Lima¹.

¹Institute of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio Grande do Norte, Natal 59072-970, RN-Brazil

²Physical Therapy Department, Federal University of Alfernas, Alfernas, 37130-000, MG, Brazil

³Department of Physiotherapy, Federal University of Rio Grande do Norte, Natal 59072-970, RN-Brazil

*** Correspondence to:** Prof. Dr. Kássio M.G. Lima, Institute of Chemistry, Biological Chemistry and Chemometrics, UFRN, Natal, 59072-970, Brazil; Email: kassiolima@gmail.com; Tel.: +55(84)3342 2323

Resumo

A saúde do idoso é um problema de saúde pública normalmente ignorada pelo governo, no qual há omissão de cuidados na maioria dos casos. A perda progressiva de massa e força muscular, clinicamente definida como sarcopenia, é um exemplo de doença que normalmente não é investigada em atendimentos primários de saúde. Hoje em dia, a sarcopenia vem se tornando uma doença muito comum entre idosos, impossibilitando o seu portador de viver de forma autônoma e o impedindo de realizar tarefas básicas como a manutenção da higiene pessoal. Neste estudo, utilizamos espectroscopia infravermelha com transformada de Fourier e a reflexão total atenuada (ATR-FTIR) juntamente com técnicas de análise multivariada para identificar e discriminar idosos portadores de sarcopenia de idosos saudáveis com o objetivo de desenvolver uma metodologia rápida, acessível, não invasiva focada na triagem dessa doença, como também a identificação de supostos biomarcadores. Técnicas como análise de componentes principais - análise discriminante linear (PCA-LDA), algoritmo de projeção sucessiva - análise discriminante linear (SPA-LDA) e algoritmo genético - análise discriminante linear (GA-LDA) foi aplicada. Um total de 174 amostras (92 saudáveis e 82 sarcopênicos) de plasma sanguíneo foi analisado por espectroscopia de ATR-FTIR. Os resultados mostraram que os modelos foram bem sucedidos na classificação de controle e sarcopênicos. Os resultados da classificação multivariada foram testados com base na sensibilidade e especificidade alcançando taxas de acerto acima de 60%, sugerindo que a espectroscopia ATR-FTIR, juntamente com algoritmos de análise multivariada, é uma ferramenta potencialmente efetiva para a triagem de sarcopenia em plasma sanguíneo de idosos.

Palavras-chave: ATR-FTIR, Sarcopenia, Idosos, Classificação Multivariada.

Abstract

Elderly health is a public health problem usually ignored by the government, in which care is often neglected. Progressive loss of muscle mass and strength, clinically defined as sarcopenia, is an example of a disease that is not normally investigated in primary health care settings. Nowadays, sarcopenia has become a very common disease among the elderly, making it impossible for the patient to live autonomously and preventing them from performing basic tasks such as maintaining personal hygiene. In this study, we used infrared spectroscopy with Fourier transform and attenuated total reflection (ATR-FTIR) along with multivariate analysis techniques to identify and discriminate elderly patients with healthy elderly sarcopenia with the objective of developing a fast, accessible, non-invasive focused methodology in screening this disease, as well as the identification of supposed biomarkers. Techniques such as main component analysis - linear discriminant analysis (PCA-LDA), successive projection algorithm - linear discriminant analysis (SPA-LDA) and genetic algorithm - linear discriminant analysis (GA-LDA) were applied. A total of 174 samples (92 healthy and 82 sarcopenic) of blood plasma were analyzed by ATR-FTIR spectroscopy. The results showed that the models were successful in classifying the disease. The results of the multivariate classification were tested based on sensitivity and specificity reaching set-up rates above 60%, suggesting that ATR-FTIR spectroscopy, together with multivariate analysis algorithms, is a potentially effective tool for the screening of blood plasma sarcopenia of the elderly.

Key words: ATR-FTIR, Sarcopenia, Elderly, Multivariate Classification.

Introdução

A perda progressiva de massa muscular é uma das características mais marcantes do processo de envelhecimento ^[1, 2], podendo conduzir à perda de força muscular e funcionalidade ^[3, 4]. Tal condição é definida como sarcopenia e tem sido considerada uma nova síndrome geriátrica, devido a sua alta prevalência e consequências deletérias para as pessoas idosas ^[3, 5].

Segundo a Organização Mundial de Saúde (OMS), nas próximas décadas a população mundial com mais de 60 anos vai passar dos atuais 841 milhões para 2 bilhões até 2050, tornando as doenças crônicas e o bem-estar da terceira idade novos desafios de saúde pública global, sendo necessárias estratégias para melhorar a prevenção e o gerenciamento de condições crônicas, disponibilizando cuidados de excelência acessíveis a todos os idosos, levando em consideração o ambiente físico e social ^[6]. A rápida identificação da sarcopenia nos serviços primários de saúde e implementação de tratamento precoce pode reduzir o desenvolvimento da incapacidade e hospitalização de idosos ^[7].

A sarcopenia tem efeitos adversos que se reforçam mutuamente, levando a dependência, maior uso de serviços de saúde, institucionalização, perda da qualidade de vida ^[4, 7]. Estudos apontam esta condição de saúde como preditora de limitações nas atividades de vida diária ^[8], desenvolvimento de incapacidade ^[9], ocorrência de quedas ^[10], fragilidade ^[5], incidência de fraturas ^[11] e mortalidade ^[12]. Sua importância no contexto do envelhecimento levou ao seu reconhecimento como uma entidade clínica, com concessão de um código na International Statistical Classification of Diseases and Related Health Problems (CID-10) (M62.84) ^[13].

O diagnóstico da sarcopenia envolve a mensuração da composição corporal, juntamente com medidas de força muscular e de performance física ^[3]. Esse diagnóstico foi validado por diferentes estudos prospectivos, apesar da variabilidade de pontos de corte nessas medidas, em consequência de variações étnicas e geográficas ^[14, 15]. Embora diferentes instrumentos tenham sido propostos para avaliar a presença de sarcopenia, considerando aspectos relacionados à força muscular e função física, clinicamente, as medidas de massa magra são consideradas como determinantes no diagnóstico da sarcopenia, pois é possível verificar se houve redução significativa da composição muscular característica da doença ^[3, 16].

Exames de imagem, como a ressonância magnética (RM) e a tomografia computadorizada (TC) são considerados padrão ouro, pois diferencia de forma precisa a massa magra do tecido gorduroso. No entanto, ambas as técnicas têm custo muito elevado e expõem os indivíduos a altos índices de radiação, dificultando o uso no âmbito clínico. A absorptometria por emissão de raios X de dupla energia (dual energy x-ray absorptiometry - DEXA) tem sido uma alternativa em relação à RM e TC, por ser uma técnica com baixa exposição à radiação, curto tempo de exame e alta precisão. A DEXA estima a massa muscular esquelética apendicular de membros superiores e inferiores a partir da mensuração dos três componentes corporais: massa de gordura, massa magra e massa óssea ^[17]. Embora tenha menor custo comparado a RM e TC, constitui ainda uma técnica onerosa para o sistema de saúde.

Medidas duplamente indiretas da composição corporal também têm sido propostas, como a bioimpedância (BIA), que consegue estimar o volume de gordura e massa magra a um baixo custo ^[18] e a circunferência de panturrilha ^[19]. Contudo, essas medidas são consideradas pouco confiáveis e recomendáveis para o diagnóstico da sarcopenia ^[3, 20]. Outra possibilidade são as equações preditivas da massa magra, usando medidas antropométricas e variáveis de sexo e raça, demonstrando bom valor preditivo para a baixa massa muscular esquelética utilizada, sendo essa a principal ferramenta clínica utilizada pela maioria dos especialistas ^[21 – 23].

Em serviços primários de saúde, um *screening* que identifique a sarcopenia é fundamental, para que após essa triagem inicial o diagnóstico seja realizado em ambiente clínico, com técnicas mais acuradas/específicas, reduzindo os custos para o sistema de saúde. Todavia, atualmente, não há disponível ferramenta, com propriedades psicométricas adequadas, para triar a sarcopenia em idosos da comunidade, em nível da atenção primária. Nesta perspectiva, verifica-se urgência em melhorar métodos de detecção atuais de sarcopenia, mais rápidos e econômicos, considerando os avanços na ciência e tecnologia, como a bioespectroscopia e análise multivariada.

A espectroscopia do infravermelho com transformada de Fourier (FTIR) é sensível ao comportamento vibracional de determinados compostos frente à radiação eletromagnética e sua aplicação tem sido avaliada como uma forte tecnologia para pesquisas biológicas ^[24, 25], onde é atualmente reconhecida como bioespectroscopia. A faixa do infravermelho médio (IVM) tem apresentado a melhor resolução quando comparada a outras regiões do espectro, uma vez que pode ser diretamente relacionada com biomoléculas tais como proteínas, lipídios, carboidratos, DNA e RNA. A região do

IVM compreende a faixa entre 400 cm^{-1} - 4000 cm^{-1} do espectro eletromagnético e essa radiação são absorvidos por biomoléculas [26].

A faixa espectral entre 900 cm^{-1} a 1800 cm^{-1} é conhecida como impressão digital biológica [27], é nessa região que as informações relacionadas às biomoléculas podem ser observadas e variações metabolômicas, proteômicas, lipidômicas e genômicas são percebidas em números de ondas característicos como lipídeos (1750 cm^{-1}), carboidratos (1155 cm^{-1}), proteínas (amida I - 1650 cm^{-1} , amida II, 1550 cm^{-1} , amida III, 1260 cm^{-1}), DNA/RNA (1225 cm^{-1} , 1080 cm^{-1}) [25, 26]. Essa metodologia não destrói a amostra, podendo a mesma ser reutilizada para outros experimentos, a obtenção do espectro é rápida e reprodutível.

Devido à complexidade dos espectros FT-IR e sobreposição espectral se faz necessário o uso de métodos quimiométricos de classificação multivariada para segregação das classes, o que garante uma maior e melhor interpretação dos dados. Ferramentas computacionais como análise de componentes principais - análise de discriminante linear (PCA-LDA), algoritmos de projeções sucessivas - análise de discriminante linear (SPA-LDA) e algoritmos genéticos - análise de discriminante linear (GA-LDA) é comprovadamente eficiente para classificar diferentes processos biológicos complexos, como descritos por THEOPHILOU et al 2015 que classificou indivíduos saudáveis e com câncer de ovário em estágio inicial e avançado utilizando o plasma sanguíneo, encontrando proteínas e lipídios biomarcadores relacionados com o desenvolvimento da doença.

Neste artigo, tivemos como objetivo a aplicação da espectroscopia ATR-FTIR e classificação multivariada para desenvolver um método de triagem de idosos com e sem sarcopenia a partir de plasma sanguíneo. Considerando a sarcopenia como um problema crônico em potencial para futuras gerações da 3ª idade, esta metodologia pode contribuir para o diagnóstico precoce de idosos com sarcopenia para prevenção de suas consequências negativas melhorando assim o bem estar e envelhecimento saudável das gerações futuras.

Materiais e Métodos.

Comitê de ética

Este estudo foi aprovado pelo Comitê de Ética para pesquisa institucional da Federal de Alfenas (UNIFAL-MG), ETIC 564.941 e conduzido de acordo com a Declaração de Helsinki. Além disso, todos os participantes forneceram consentimento informado por escrito antes de participar deste estudo. Trata-se de um estudo transversal com duração de agosto de 2015 a junho de 2016

Seleção dos participantes

A população total do estudo foi composta por idosos com 60 anos ou mais, residente nas áreas de abrangência da Estratégia de Saúde da Família (ESF) do município de Alfenas-MG, Brasil.

A população total do estudo foi composta por 521 idosos (acima de 60 anos) residentes na área de abrangência da Estratégia de Saúde da Família (ESF) do município de Alfenas-MG. Os idosos foram selecionados aleatoriamente a partir da lista de usuários cadastrados na ESF considerando $\alpha=0,05$ e *power* de 80%. A amostragem considerou a distribuição de idosos cadastrados, randomizando o número de idosos de forma proporcional a cada uma das regiões de abrangência do serviço primário de saúde do município, garantindo representatividade populacional.

Como critérios de exclusão foram utilizados: idosos com alterações cognitivas detectáveis pelo Mini-Exame do Estado Mental (MEEM) ^[29], doença inflamatória ou infecciosa em fase aguda; neoplasia nos últimos cinco anos; uso de drogas imunossupressoras; amputações nos membros inferiores; cirurgias ou fraturas nos membros inferiores nos últimos seis meses; presença de doenças ou sequelas neurológicas que impedissem a realização dos testes propostos.

Para o presente estudo, foi utilizado como critério de triagem dos idosos à presença de sarcopenia segundo o algoritmo proposto pelo European Working Group on Sarcopenia in Older people (EWGSOP) ^[3], em que idosos com baixa massa muscular associada à força reduzida de prensão palmar e/ou função física (velocidade de marcha) foram consideradas sarcopênicos. Em seguida, os idosos considerados com sarcopenia foram pareados àqueles sem sarcopenia, segundo a idade e sexo.

Obtenção de dados biométricos e biológicos

As entrevistas e testes funcionais foram realizados nas residências dos idosos, enquanto a coleta de sangue foi realizada no laboratório Central de Análises Clínicas da Universidade Federal de Alfenas- MG.

Biometria: Para caracterização da amostra foram consideradas as seguintes variáveis: idade, sexo, raça, escolaridade, renda familiar, estado civil e medidas antropométricas como peso (kg) e altura (m). A massa apendicular esquelética (MAE) foi estimada a partir da equação de Lee como segue: $MAE = (0,244 \times \text{peso}) + (7,8 \times \text{altura}) + (6,6 \times \text{sexo}) - (0,098 \times \text{idade}) + (\text{raça} - 3,3)$, onde o valor 0 foi assumido para mulheres e 1 para homens, 0 para brancos, 1,4 para afro-americanos e -1,2 para asiáticos. O peso mensurado em quilogramas e altura em metros. Esta fórmula é preditiva da massa muscular e foi validado para a população idosa brasileira com excelentes valores de correlação com a DEXA, instrumento considerado padrão-ouro para avaliação da massa magra ($r=0,86$ para homens e $r= 0,90$ para mulheres). A concordância entre a densitometria e a equação para determinar a prevalência de sarcopenia foi forte ($k=0.74$; $p < 0,001$) com alta especificidade (89%) e sensibilidade (86%).^[22] Os valores encontrados de MAE foram então ajustados de acordo com o quadrado da altura para estabelecer o índice de massa muscular.

Realização de testes funcionais: A função física foi avaliada pela velocidade de marcha, em um percurso de 4 metros, em uma superfície plana. O tempo para percorrer a distância foi mensurado com auxílio de um cronômetro. A velocidade de marcha foi calculada em metros por segundo (m/s). O ponto de corte igual ou menor a 0,8 m/s foi considerado como positivo para a identificação do idoso com sarcopenia segundo o algoritmo proposto pelo EWGSOP.^[3]

A força de preensão palmar foi mensurada por meio do dinamômetro JAMAR[®] (modelo PC5030II), usando o protocolo recomendado pela *The American Society of Hand Therapists*. Para realizar as medidas os idosos foram posicionados sentados, de forma confortável, em uma cadeira sem apoio para os braços, com os ombros em posição neutra, o cotovelo do membro dominante a ser medido era mantido flexionado em 90 graus, e o antebraço em rotação neutra^[30]. A mensuração foi realizada com a mão dominante, sendo obtidas três medidas em quilogramas força (Kgf), com intervalo de um minuto entre cada execução. Foi considerada a média das três medidas. O

instrumento apresenta alta confiabilidade para a população idosa (ICC >0.90).^[31] Os pontos de corte para determinar fraqueza muscular adotado foram < 16 kgf para mulheres e < 26 kgf para homens.^[32]

Coleta de sangue: A coleta das amostras de sangue foi realizada em tubos a vácuo, com anticoagulante EDTA. As amostras foram centrifugadas em 4000 RPM, por 10 minutos, e o plasma removido em ambiente estéril, dentro de 30 minutos após a coleta, e as alíquotas estocadas em freezer -80°C até as análises.

Obtenção dos espectros ATR – FTIR

As medições espectrais foram realizadas utilizando um espectrofotômetro de marca SHIMADZU, modelo IRAffinity-1, equipado com o acessório ATR. Os espectros ATR-FTIR foram obtidos numa faixa de 700 cm⁻¹ a 4000 cm⁻¹, com uma resolução espectral de 4 cm⁻¹, a medição do tempo foi de 34 s (32 varreduras) por espectro. O ar foi utilizado com branco para todas as amostras. Uma gota de plasma foi adicionada diretamente sobre o cristal de ATR e um espectro foi obtido, em seguida o cristal foi limpo com álcool 70% e outra gota do mesmo plasma foi adicionada e obtida um novo espectro. Esse procedimento foi realizado em triplicata dando um total de 522 espectros ATR-FTIR sendo 276 controles e 246 sarcopênicos. As medições espectrais foram feitas com temperatura controlada de 22 ° C e 60% de umidade relativa do ar.

Análise computacional

A importação dos dados, pré-tratamento e construção dos modelos de classificação multivariada (PCA-LDA, SPA-LDA e GA-LDA) foram desenvolvidos no software MATLAB R2014a (<http://www.mathworks.com>). Foi obtida a média por amostra (n=174 no total, sendo 92 controles e 82 sarcopênicos) dos espectros brutos e pré-processados cortando de 1300 a 1750 cm⁻¹, aplicando filtro de savitzky-golay com janela de 15 pontos e MSC (*Multiplicative Scatter Correction*).

A análise exploratória de dados foi realizada a fim de observar a distribuição geral dos dados e se haveria agrupamento das classes controle e sarcopênicos somente com o uso de uma ferramenta não supervisionada. A redução do número de variáveis através da aplicação da PCA foi alcançada através da decomposição da matriz original X em seus produtos matriciais em *scores* e *loadings*, como mostrado na equação (1):

$$X = TP^T + E \quad (1)$$

Onde: X é a matriz $I \times J$ (I é o número de amostras e J é o número de variáveis); T é a matriz de escores $I \times A$ (A é o número de componentes calculados); P é a matriz de loadings $J \times A$ e E é a matriz residual $I \times J$.

Para os modelos PCA-LDA, SPA-LDA e GA-LDA as amostras foram divididas em um conjunto treinamento (70%), um conjunto validação (15%) e um conjunto previsão (15%) aplicando nos espectros IVM o clássico algoritmo de amostragem uniforme Kennard-Stone (KS). O algoritmo KS foi aplicado separadamente em cada classe a fim de extrair a maior representatividade de cada classe e maximizar a amostragem do conjunto de dados. As amostras de treinamento foram utilizadas no processo de modelagem como também a seleção de variáveis para dos modelos LDA enquanto que o conjunto de previsão somente foi utilizado na etapa final da classificação. O número de variáveis ótimas selecionadas pelo SPA-LDA e GA-LDA foram calculadas com a função de custo G e é dado por:

$$G = \frac{1}{N_v} \sum_{n=1}^{N_v} g_n \quad (2)$$

onde g_n é definido como

$$g_n = \frac{r^2(x_n, m_{l(n)})}{\min_{l(m) \neq l(n)} r^2(x_n, m_{l(n)})} \quad (3)$$

sendo $l(n)$ é o índice da classe verdadeira para o n ésimo objeto da validação $X_{(n)}$.

A rotina do GA-LDA foi trabalhada durante 24 gerações com 48 cromossomos cada. As probabilidades de cruzamento e mutação foram ajustadas para 60% e 10% respectivamente. A melhor solução resultante das três realizações do GA foi mantida constante. Para esse estudo, os valores da função discriminante do LDA foram obtidos por classe de tratamento. A Sensibilidade (SENS) corresponde à certeza de que uma amostra pertence realmente a uma das categorias de classe e especificidade (SPEC) é a certeza de que uma amostra não pertence a uma das categorias de classe. Esses parâmetros medem a qualidade dos modelos calculados e são utilizados nesse estudo para avaliar a taxa de acerto na classificação de indivíduos controle e com sarcopenia. Podemos calcular usando as seguintes equações:

$$\text{SENS}(\%) = \frac{TP}{TP+FN} \times 100 \quad (4)$$

$$ESPC(\%) = \frac{TN}{TN+FP} \times 100 \quad (5)$$

Onde, TP é definido como o verdadeiro positivo, TN é o verdadeiro negativo, FP é o falso positivo e FN é o falso negativo.

Resultados e discussão

Dos 521 idosos avaliados, a amostra de plasma foi disponível para 406. Destes, 82 idosos foram identificados com sarcopenia de acordo com o algoritmo proposto pelo EWGSOP e pareados com idosos sem sarcopenia de acordo com sexo e idade. As características sociodemográficas e clínicas da amostra são apresentadas na tabela 1.

Tabela 1: Características descritivas médias sociodemográficas e clínicas dos idosos voluntários diagnosticados com e sem sarcopenia pelo algoritmo proposto pelo EWGSOP.

Variáveis	Sarcopenia (n=61)	Sem Sarcopenia (n=61)	p-valor
	Media (SD) ou N (%)	Media (SD) ou N (%)	
Idade	74.42 (7.95)	73.86 (7.23)	0.05
Sexo			
Homens	19 (31.3)	19 (31.3)	1.00
Mulheres	42 (68.9)	42 (68.9)	
Anos de estudo	2.45 (2.36)	3.80 (4.01)	0.74
Suficiência de renda			
Suficiente	39 (63.9)	31 (50.8)	0.14
Insuficiente	30 (49.2)	22 (36.1)	
Saúde auto relatada			
Boa	31 (50.8)	29 (47.5)	0.58
Moderada	29 (47.5)	29 (47.5)	
Ruim	1 (1.6)	3 (4.9)	
Peso (kg)	55.22 (10.39)	72.75 (11.66)	0.06
Altura (m)	1.57 (0.07)	1.59 (0.09)	<0.01
IMC (kg/m ²)	21.54 (2.94)	29.01 (4.85)	0.13
Massa muscular esquelética	7.00 (1.45)	8.71 (1.47)	<0.01
Velocidade da marcha (m/s)	0.56 (0.17)	0.65 (0.23)	0.14
Força de preensão palmar (kg)	22.95 (9.18)	24.80 (8.81)	<0.01

A média dos espectros brutos e após o pré-processamento para cada classe pode ser observado na Fig. 1.

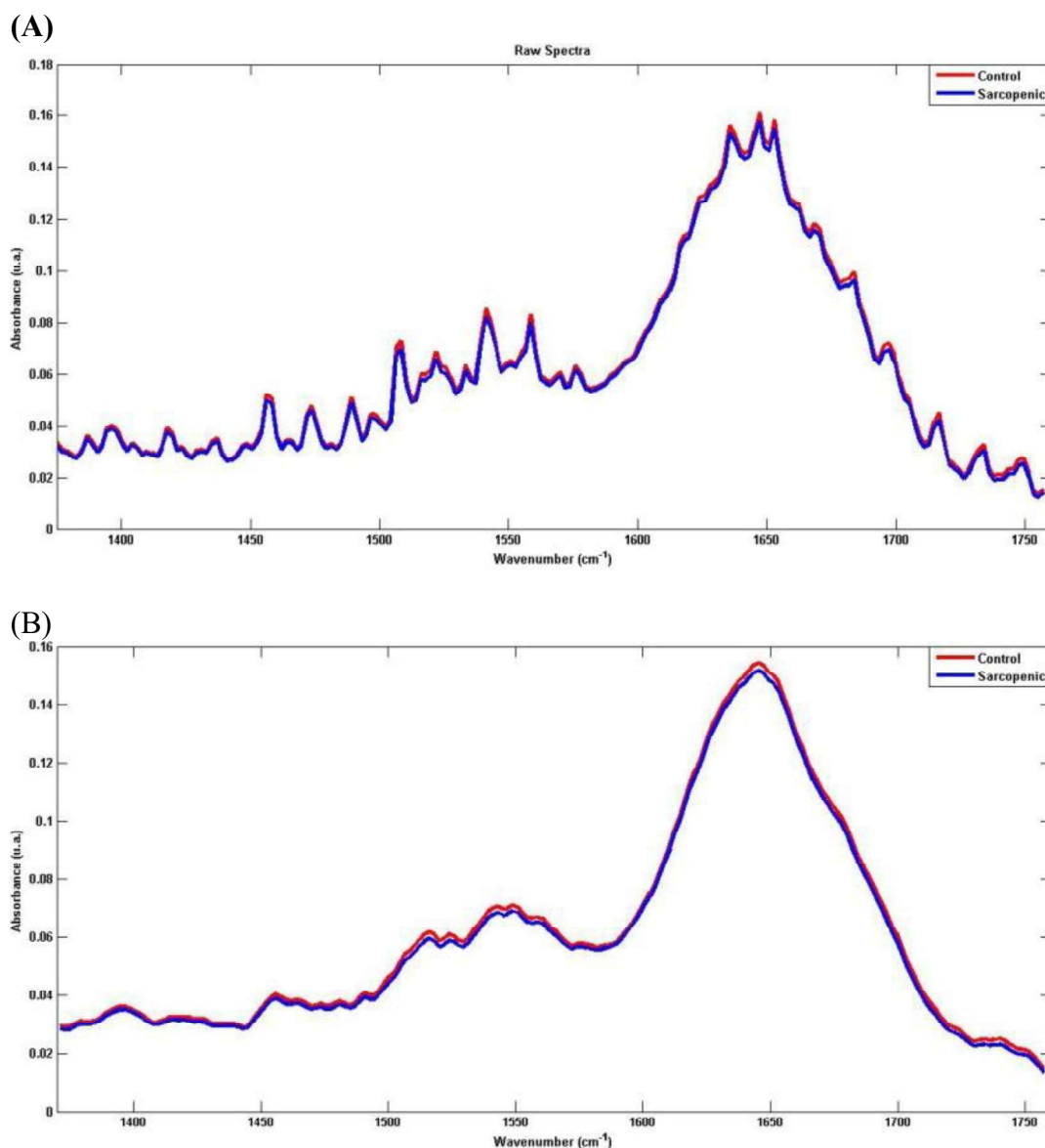


Figura 1: (A) Espectros brutos para cada categoria. (B) Espectros pré-processados para cada categoria. O espectro controle é mostrado em vermelho; O espectro do grupo com sarcopenia é mostrado em azul.

O conjunto de espectros apresenta elevada semelhança entre si, tornando impossível a percepção das diferenças através da observação visual do espectro. Dessa forma, foi necessária a utilização de uma ferramenta multivariada para maximizar as possíveis diferenças bioquímicas que melhor caracterizam cada classe. Para este conjunto foi utilizada a Análise de Componentes Principais (PCA). Após aplicação da PCA, um total de 3 componentes principais explicou 98% da variabilidade dos dados e não houve possibilidade de se observar uma separação evidente das classes controle e sarcopênico, como pode ser observado na Fig. 2.

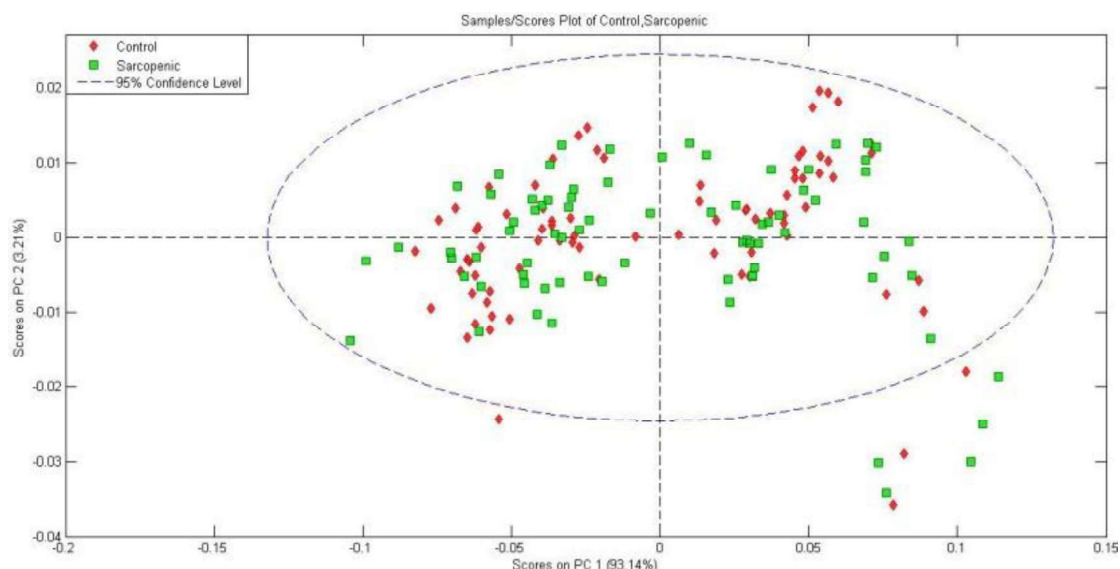
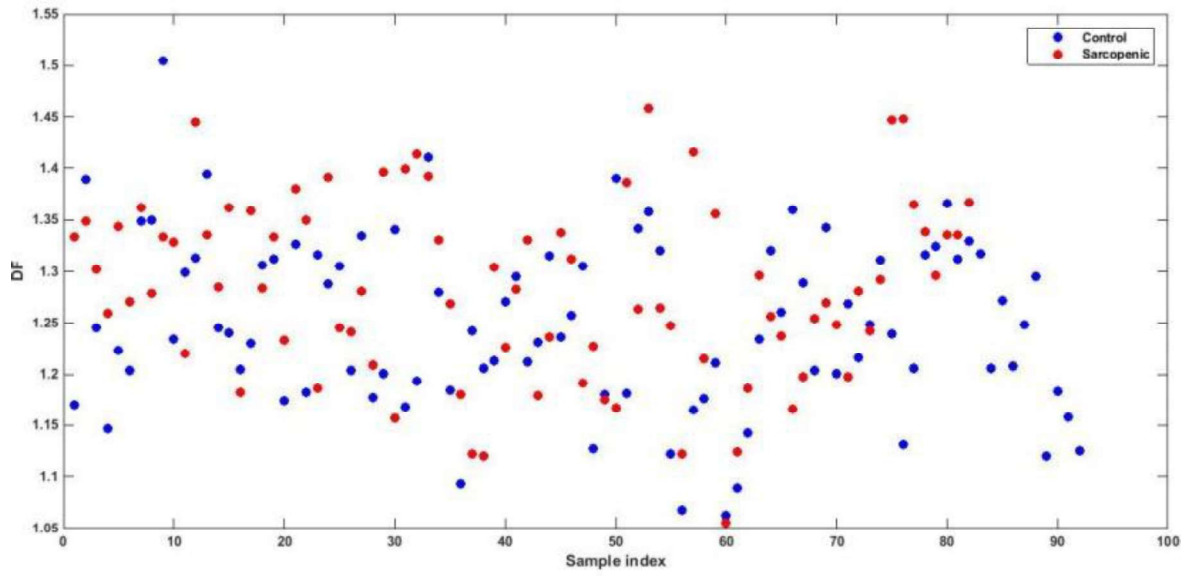


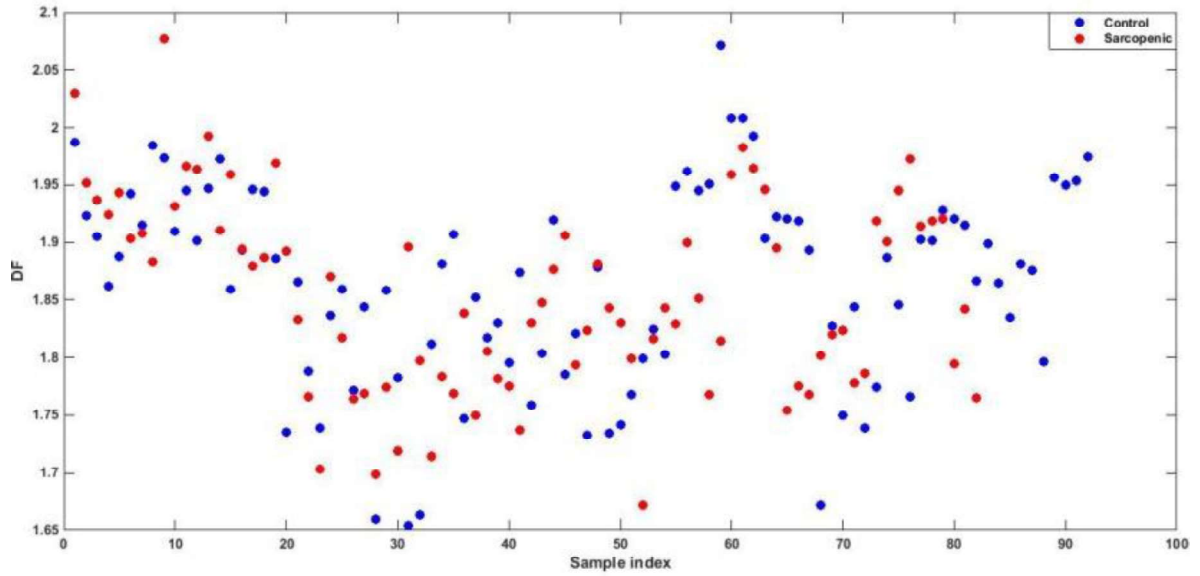
Figura 2: A aplicação da análise exploratória por PCA para segregação de controle (vermelho) e sarcopênico (verde) . *Scores* de PC1xPC2.

Portanto, é necessária a aplicação de métodos de classificação supervisionada com capacidade de redução de dimensionalidade (PCA) e seleção de variáveis (SPA e GA) juntamente com a LDA com o objetivo de obter a melhor segregação entre indivíduos controle e com sarcopenia. A Fig. 3A mostra os escores da PCA-LDA (DF1 x Amostras) para as classes controle e sarcopênico, mostrando que existe uma separação entre as classes com uma aparente região de fronteira. Para o SPA-LDA usando dois números de onda selecionados (respectivamente, 1610 cm^{-1} e 1695 cm^{-1}), obtidos pela função de custo G, alcançando uma segregação melhorada entre controle e sarcopênicos (Fig. 3B) quando comparada com PCA-LDA. No entanto, quando GA-LDA foi empregado aplicando uma classificação evolutiva, usando 4 números de onda selecionados (respectivamente, 1479 cm^{-1} , 1620 cm^{-1} , 1622 cm^{-1} e 1699 cm^{-1}) proporcionaram uma maior segregação comparada com PCA-LDA e SPA-LDA em conjunto, como mostrado na Fig. 3C.

(A)



(B)



(C)

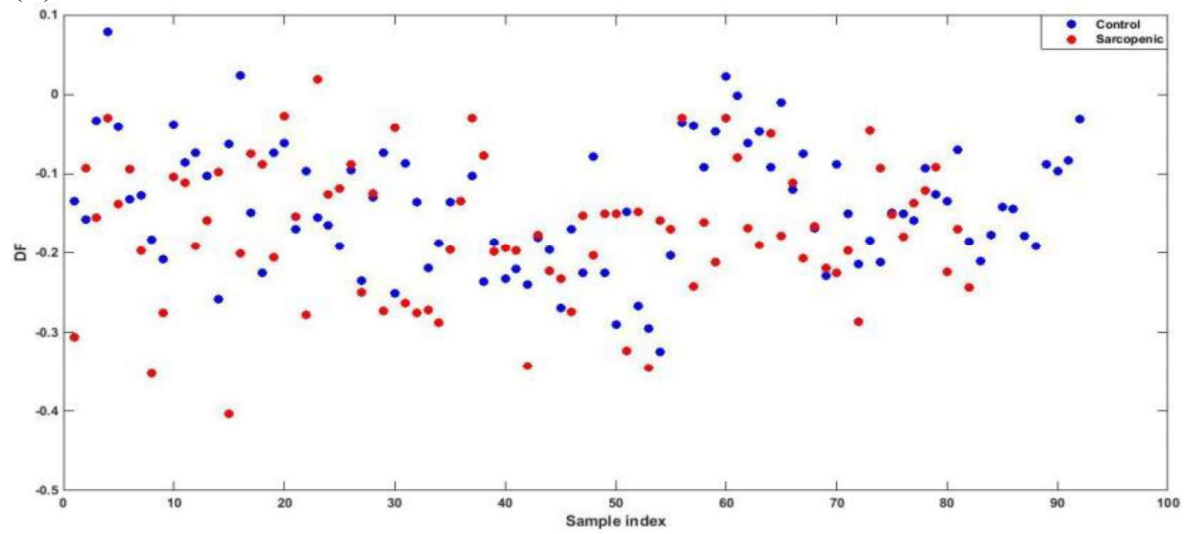
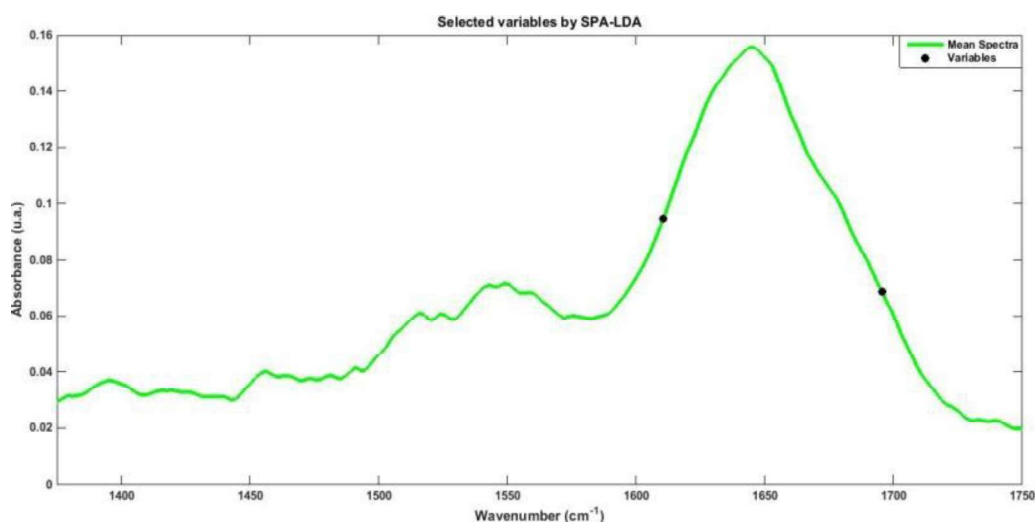


Figure 3: A aplicação da análise discriminante linear à categoria para segregação de controle (vermelho) e sarcopênico (azul). (A) DF1 x amostras da função discriminante calculada usando análise de componentes principais (PCA) - análise discriminante linear (LDA) de cada categoria. (B) DF1 x amostras da função discriminante calculada usando as variáveis selecionadas por algoritmos de projeção sucessivos (SPA) - análise discriminante linear (LDA) de cada categoria. (C) DF1 x amostras da função discriminante calculados usando as variáveis selecionadas por algoritmos genéticos (GA) - análise discriminante linear (LDA) de cada categoria.

A Fig. 4 demonstra as variáveis selecionadas pelo SPA-LDA e GA-LDA que diferem em valores, no entanto representa a mesma região do espectro. O exame dos números de onda selecionados após a modelagem computacional mostrou que as principais alterações bioquímicas discriminavam indivíduos controle de indivíduos sarcopênicos. Vários números de onda selecionados parecem ser de particular interesse e já relatados na literatura como biomarcadores para diagnóstico de sarcopenia em idosos, ou seja, as variáveis 1479 cm^{-1} , 1610 cm^{-1} , 1620 cm^{-1} , 1622 cm^{-1} , 1695 cm^{-1} e 1699 cm^{-1} estão associados à banda vibracional de proteínas (uma combinação de banda de amida I e amida II de polipeptídeos).

(A)



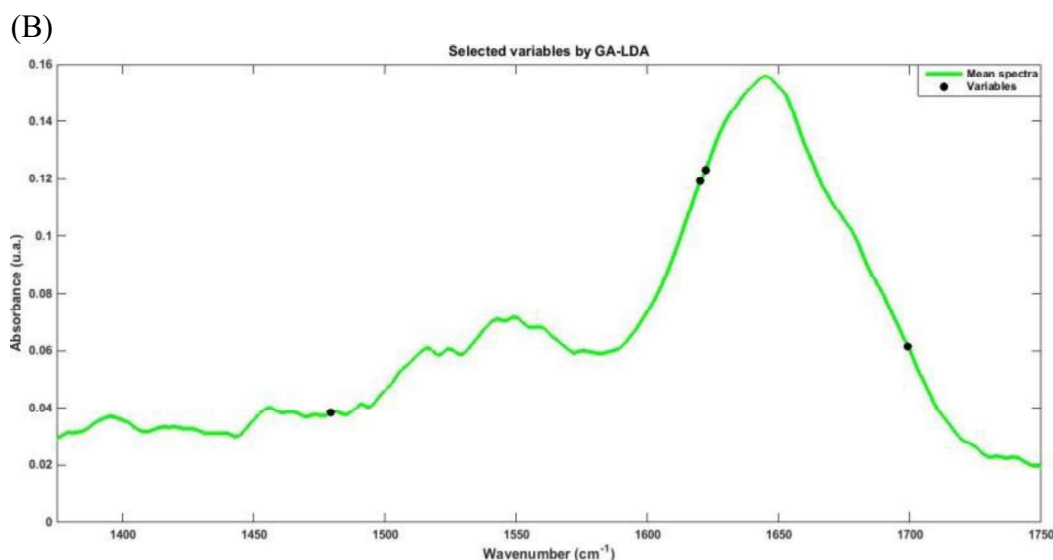


Figure 4: Espectro médio da região de impressão digital biológico com as variáveis selecionadas por: (A) SPA-LDA e (B) GA-LDA que resultaram nas melhores classificações.

As taxas de classificação foram determinadas usando os melhores modelos. Na Tabela 2 estão os resultados de validação das categorias controle e sarcopênicos.

Table 2: Medida de desempenho para PCA-LDA, SPA-LDA e GA-LDA na classificação de Controle e indivíduos Sarcopênicos por biospectroscopia ATR-FTIR.

		Control		Sarcopenic
PCA - LDA	sens	46.7	sens	70.0
	spec	66.7	spec	50.0
SPA - LDA	sens	86.7	sens	30.0
	spec	40.0	spec	90.0
GA - LDA	sens	66.7	sens	40.0
	spec	46.7	spec	50.0

Além disso, os valores de sensibilidade e especificidade demonstram que utilizando a região de impressão digital biológico, a distinção entre categorias de controle e sarcopênicos utilizando os algoritmos PCA-LDA, SPA-LDA e GA-LDA é possível e indicando que a biospectroscopia ATR-FTIR em conjunto com quimiometria é uma poderosa ferramenta para identificar possíveis biomarcadores moleculares e classificar potenciais pacientes portadores de sarcopenia.

Conclusões

Concluimos que o uso conjunto da bioespectroscopia ATR-FTIR e classificação multivariada com análise discriminante linear é uma ferramenta em potencial para a identificação de indivíduos com sarcopenia e indivíduos saudáveis utilizando o plasma sanguíneo. As técnicas de seleção de variáveis (SPA e GA) usando LDA foram realizadas em uma tentativa de obter mais informações sobre as possíveis diferenças entre as categorias de espectro ATR-FTIR, como sinalizar ligações químicas possíveis de serem correlacionadas com grupos funcionais presentes em importantes biomoléculas como proteínas, considerados forte candidatos à biomarcadores moleculares capazes de diferenciar as classes. Sendo assim, o presente estudo indica que a bioespectroscopia ATR-FTIR suportada pela classificação multivariada obteve taxas de acerto semelhantes aos métodos clínicos já utilizados no dia a dia, podendo oferecer potencial para diagnóstico em um futuro próximo. Este estudo foi realizado em fase piloto, então novas amostras devem ser incorporadas aos modelos apresentados para que haja contemplação da maior variabilidade possível e consequentemente aumento das taxas de acerto. A validação do método ocorrerá posteriormente comparando simultaneamente os resultados da classificação multivariada com os resultados da DEXA.

Agradecimentos

Leomir Aires Silva de Lima gostaria de agradecer a PPGQ-UFRN e a Capes pelo apoio científico e financeiro. K.M.G. Lima reconhece o projeto CNPq / Capes (Grant 070/2012 e 442087 / 2014-4), FAPERN (PPP 005/2012) para apoio financeiro.

Referências

- [1] Rosenberg, I. H. (1997). Sarcopenia: origins and clinical relevance. *The Journal of Nutrition*, 127(5 Suppl), 990S–991S. Retrieved from <http://jn.nutrition.org/content/127/5/990S.full>
- [2] Sayer, A. A., Syddall, H., Martin, H., Patel, H., Baylis, D., & Cooper, C. (2008). The developmental origins of sarcopenia. *Journal of Nutrition, Health and Aging*, 12(7), 427. <https://doi.org/10.1007/BF02982703>
- [3] Cruz-Jentoft, A. J., Baeyens, J. P., Bauer, J. M., Boirie, Y., Cederholm, T., Landi, F., Zamboni, M. (2010). Sarcopenia: European consensus on definition and diagnosis. *Age and Ageing*, 39(4), 412–423. <https://doi.org/10.1093/ageing/afq034>
- [4] Landi, F., Liperoti, R., Fusco, D., Mastropaolo, S., Quattrocioni, D., Proia, A., Onder, G. (2012). Prevalence and risk factors of sarcopenia among nursing home older residents. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 67 A(1), 48–55. <https://doi.org/10.1093/gerona/qlr035>
- [5] Morley, J. E., von Haehling, S., Anker, S. D., & Vellas, B. (2014). From sarcopenia to frailty: A road less traveled. *Journal of Cachexia, Sarcopenia and Muscle*, 5(1), 5–8. <https://doi.org/10.1007/s13539-014-0132-3>
- [6] Integrated care for older people: guidelines on community-level interventions to manage declines in intrinsic capacity. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO

- [7] Morley, J. E., & Cao, L. (2015). Rapid screening for sarcopenia. *Journal of Cachexia, Sarcopenia and Muscle*, 6(4), 312–314. <https://doi.org/10.1002/jcsm.12079>
- [8] Al Snih, S., Markides, K., Ottenbacher, K., & Raji, M. (2004). Hand grip strength and incident ADL disability in elderly Mexican Americans over a seven-year period. *Aging Clinical and Experimental Research*, 16(6), 481–486. <https://doi.org/10.1007/BF03327406>
- [9] Hirani, V., Blyth, F., Naganathan, V., Le Couteur, D. G., Seibel, M. J., Waite, L. M., Cumming, R. G. (2015). Sarcopenia Is Associated With Incident Disability, Institutionalization, and Mortality in Community-Dwelling Older Men: The Concord Health and Ageing in Men Project. *Journal of the American Medical Directors Association*, 16(7), 607–613. <https://doi.org/10.1016/j.jamda.2015.02.006>
- [10] Landi, F., Liperoti, R., Russo, A., Giovannini, S., Tosato, M., Capoluongo, E., Onder, G. (2012). Sarcopenia as a risk factor for falls in elderly individuals: results from the iLSIRENTE study. *Clinical Nutrition (Edinburgh, Scotland)*, 31(5), 652–8. <https://doi.org/10.1016/j.clnu.2012.02.007>
- [11] Chalhoub, D., Cawthon, P. M., Ensrud, K. E., Stefanick, M. L., Kado, D. M., Boudreau, R., Cauley, J. A. (2015). Risk of nonspine fractures in older adults with sarcopenia, low bone mass, or both. *Journal of the American Geriatrics Society*, 63(9), 1733–1740. <https://doi.org/10.1111/jgs.13605>
- [12] Brown, J. C., Harhay, M. O., & Harhay, M. N. (2015). Sarcopenia and mortality among a population-based sample of community-dwelling older adults. *Journal of Cachexia, Sarcopenia and Muscle*, (October 2015), 290–298. <https://doi.org/10.1002/jcsm.12073>
- [13] Anker, S. D., Morley, J. E., & von Haehling, S. (2016). Welcome to the ICD-10 code for sarcopenia. *Journal of Cachexia, Sarcopenia and Muscle*, 7(5), 512–514. <https://doi.org/10.1002/jcsm.12147>
- [14] Bianchi, L., Ferrucci, L., Cherubini, A., Maggio, M., Bandinelli, S., Savino, E.,

Volpato, S. (2015). The Predictive Value of the EWGSOP Definition of Sarcopenia: Results From the InCHIANTI Study. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 71(2), glv129-. <https://doi.org/10.1093/gerona/glv129>

[15] Woo, J., Arai, H., Ng, T. P., Sayer, A. A., Wong, M., Syddall, H., ... Zhang, T. M. (2014). Ethnic and geographic variations in muscle mass, muscle strength and physical performance measures. *European Geriatric Medicine*, 5(3), 155–164. <https://doi.org/10.1016/j.eurger.2014.04.003>

[16] Mijnders, D. M., Meijers, J. M. M., Halfens, R. J. G., ter Borg, S., Luiking, Y. C., Verlaan, S., Schols, J. M. G. a. (2013). Validity and reliability of tools to measure muscle mass, strength, and physical performance in community-dwelling older people: a systematic review. *Journal of the American Medical Directors Association*, 14(3), 170–8. <https://doi.org/10.1016/j.jamda.2012.10.009>

[17] Heymsfield, S. B., Adamek, M., Gonzalez, M. C., Jia, G., & Thomas, D. M. (2014). Assessing skeletal muscle mass: Historical overview and state of the art. *Journal of Cachexia, Sarcopenia and Muscle*, 5(1), 9–18. <https://doi.org/10.1007/s13539-014-0130-5>

[18] Goodman, M. J., Ghate, S. R., Mavros, P., Sen, S., Marcus, R. L., Joy, E., & Brixner, D. I. (2013). Development of a practical screening tool to predict low muscle mass using NHANES 1999-2004. *Journal of Cachexia, Sarcopenia and Muscle*, 4(3), 187–197. <https://doi.org/10.1007/s13539-013-0107-9>

[19] Rolland, Y., Czerwinski, S., Abellan Van Kan, G., Morley, J. E., Cesari, M., Onder, G., Vellas, B. (2008). Sarcopenia: its assessment, etiology, pathogenesis, consequences and future perspectives. *The Journal of Nutrition, Health & Aging*, 12(7), 433–50. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3988678&tool=pmcentrez&rendertype=abstract>

[20] Rolland, Y., Lauwers-Cances, V., Cournot, M., Nourhashemi, F., Reynish, W., Riviere, D., Grandjean, H. (2003). Sarcopenia, calf circumference, and physical

function of elderly women: a cross-sectional study. *Journal of the American Geriatrics Society*, 51(8), 1120–1124. <https://doi.org/10.1046/j.1532-5415.2003.51362.x>

[21] Lee, W. J., Liu, L. K., Peng, L. N., Lin, M. H., & Chen, L. K. (2013). Comparisons of Sarcopenia Defined by IWGS and EWGSOP Criteria Among Older People: Results From the I-Lan Longitudinal Aging Study. *Journal of the American Medical Directors Association*, 14(7), 528.e1-528.e7. <https://doi.org/10.1016/j.jamda.2013.03.019>

[22] Rech, C. R., Dellagrana, R. A., Marucci, M. de F. N., & Petroski, E. L. (2012). Validity of anthropometric equations for the estimation of muscle mass in the elderly. *Revista Brasileira de Cineantropometria & Desempenho Humano*, 14(1), 23–31. <https://doi.org/10.5007/1980-0037.2012v14n1p23>

[23] Yu, S., Appleton, S., Chapman, I., Adams, R., Wittert, G., Visvanathan, T., & Visvanathan, R. (2015). An Anthropometric Prediction Equation for Appendicular Skeletal Muscle Mass in Combination With a Measure of Muscle Function to Screen for Sarcopenia in Primary and Aged Care. *Journal of the American Medical Directors Association*, 16(1), 25–30. <https://doi.org/10.1016/j.jamda.2014.06.018>

[24] Heraud, P.; Tobin, M. J. The emergence of biospectroscopy in stem cell research. *Stem Cell Research*, v. 3, n. 1, p. 12–14, 2009. <https://doi.org/10.1016/j.scr.2009.04.002>.

[25] Kelly, J. G.; Trevisan, J.; Scott, A. D.; Carmichael, P. L.; Pollock, H. M.; Martin-Hirsch, P. L.; Martin, F. L. Biospectroscopy to metabolically profile biomolecular structure : a multistage approach linking computational analysis with biomarkers. *J. Proteome Res.* p. 1437–1448, 2011. <https://doi.org/10.1021/pr101067u>.

[26] Obinaju, B. E.; Martin, F. L. Novel biospectroscopy sensor technologies towards environmental health monitoring in urban environments health monitoring in urban environments. *Environmental Pollution*, v. 183, n. December 2012, p. 46–53, 2015. <https://doi.org/10.1016/j.envpol.2012.11.011>

- [27] Trevisan, J.; Angelov, P. P.; Carmichael, P. L.; Scott, A. D.; Martin, F. L. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst*, 2012, **137**, 3202-3215. <https://doi.org/10.1039/C2AN16300D>
- [28] Theophilou, G.; Lima, K. M. G.; Hirsch, P. L. M., Stringfellow, H. F.; Martin, F. L. ATR-FTIR spectroscopy coupled with chemometric analysis discriminates normal, borderline and malignant ovarian tissue: classifying subtypes of human cancer. *Analyst*, 2016, **141**, 585-594. <https://doi.org/10.1039/c5an00939a>
- [29] Bertolucci PH, Brucki SM, Campacci SR, J. Y. (1994). The Mini-Mental State Examination in a general population: impact of educational status. *Arq Neuropsiquiatr.*, 52(1), 1–7.
- [30] Crosby CA, Wehbé MA, M. B. (1994). Hand strength: normative values. *J Hand Surg Am.*, 19(4), 665–70.
- [31] Schrama, P. P. M., Stenneberg, M. S., Lucas, C., & van Trijffel, E. (2014). Intraexaminer Reliability of Hand-Held Dynamometry in the Upper Extremity: A Systematic Review. *Archives of Physical Medicine and Rehabilitation*, 95(12), 2444–2469. <https://doi.org/10.1016/j.apmr.2014.05.019>
- [32] Fernandes de Souza Barbosa, J., Ulises Perez Zepeda, M., Béland, F., Guralnik, J. M., Victoria Zunzunegui, M., Oliveira Guerra, R., Guerra, R. O. (2016). Clinically relevant weakness in diverse populations of older adults participating in the International Mobility in Aging Study. *Age (Dordrecht, Netherlands)*, 38(1), 25. <https://doi.org/10.1007/s11357-016-9888-z>

Capítulo 4

Identification using classification analysis of flunitrazepam in necrophagous larvae via differential pulse voltammetry and fluorescence excitation-emission matrix (EEM) spectroscopy.

Leomir A. S. de Lima

Camilo L. M. Morais

Renata A. Gama

Jéssica T. Jales

Sherlan G. Lemos

Kássio M. G. Lima.

Manuscrito submetido à revista Journal of the Brazilian Chemical Society.

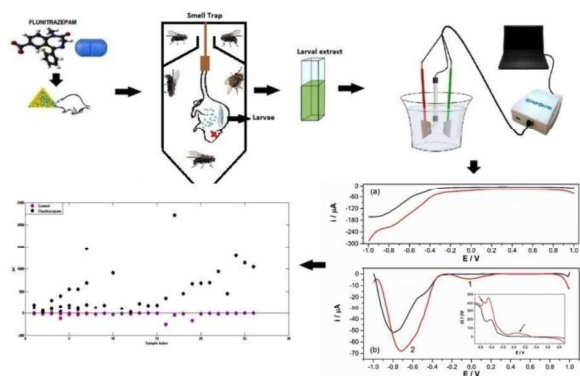
Contribuição:

- Realizei o experimento e aquisição espectral e voltamograma;
- Realizei o processamento dos dados e construção dos modelos multivariados;
- Escrevi a primeira versão do manuscrito.

Leomir A. S. de Lima

Prof. Kássio M. G. Lima

Graphical abstract



Experimental procedure to obtain samples of fly larvae extract and subsequent analysis by differential pulse voltammetry with E/V vs. $Ag/AgCl$ reference electrode and molecular fluorescence spectroscopy.

**Identification using classification analysis of flunitrazepam in
necrophagous larvae via differential pulse voltammetry and fluorescence
excitation-emission matrix (EEM) spectroscopy**

*Leomir A. S. de Lima^a, Camilo L. M. Morais^b, Jéssica T. Jales^c, Renata A. Gama^c, Sherlan
G. Lemos^d, Kássio M. G. Lima^{a*}*

^aQuímica Biológica e Quimiometria, Instituto de Química, Universidade Federal do Rio Grande do Norte, Natal, 59078-970, RN, Brasil. ^b Faculdade de Farmácia e Ciências Biomédicas, Universidade de Lancashire, Preston PR1 2HE, Reino Unido. ^c Departamento de Microbiologia e Parasitologia, Universidade Federal do Rio Grande do Norte, Natal, 59072-970, RN, Brasil. ^d Departamento de Química, Universidade Federal da Paraíba, Joao Pessoa, 58051-390, PB, Brasil.

*Email: kassiolima@gmail.com; Tel.: +55(84)3342 2323

Abstract: The use of insects to identify drugs in a cadaver has often been explored in the field of entomotoxicology. There are accurate methods; however, they require a high cost and are very time-consuming. The objective of this study was to develop two methods based on differential pulse voltammetry (DPV) and fluorescence emission excitation matrix (EEM) spectroscopy to classify necrophagous larvae (*Chrysomya megacephala*, *C. albiceps*, *Lucilia sp* and *Cochliomyia macellaria*) containing flunitrazepam. The voltammograms of larval extract samples were analyzed by principal component analysis (PCA), successive projection algorithm (SPA) and genetic algorithm (GA) - linear and quadratic discriminant analysis (LDA and QDA). The EEM fluorescence data from larval extract samples were analyzed by 2D and parallel factor analysis (PARAFAC) with linear discriminant analysis (LDA). These results suggest that DPV and EEM combined with chemometrics can be used as tools for the classification of flunitrazepam in fly larvae presenting innovative applications.

Keywords: Differential pulse voltammetry; fluorescence spectroscopy; flunitrazepam; entomotoxicology; necrophagous larvae.

Introduction

Forensic entomology is the science that investigates traces of insects and other arthropods for criminal, legal or civil investigative purposes.¹ It is often requested for estimation of minimum postmortem interval.²⁻³ However, recent techniques have allowed the collection of robust entomological evidence to resolve questions about the corpse movement, submersion interval, decapitation or dismemberment time, identification of specific traumas and artifacts postmortem cells.⁴⁻⁵ More recently, the identification of drug of abuse or other chemicals important to the crime scene was achieved in insects, which gave rise to a new research field, Entomotoxicology.⁶⁻⁷ Flunitrazepam has been reported as the rape date drug. Its illicit consumption added to alcoholic beverages strengthens its sedative effects, favoring the action of criminals, in which the victims become defenseless due to the loss of muscular control and loss of consciousness, being vulnerable to assaults and rape. Illicit use of flunitrazepam is also common in nightclubs, in which youngsters and adults use the drug individually or in combination with other drugs such as cocaine and heroin.

In skeletons or bodies in an advanced stage of decomposition, biological materials such as blood, urine or internal organs may not be available. Considering the necrobiontophagous nature of some Diptera larvae, when feeding of tissue of the corpses of intoxicated humans, they introduce in their own metabolism toxins or drugs used by the individuals when in life. Thus, by receiving target substances from the human organism, some insects, especially fly larvae, can serve as an alternative for toxicological analysis.⁸

In addition to the identification of drugs and toxins in corpses and the development of immature insects, entomotoxicology may be useful in investigating the effects of such

substances on the development of arthropods in order to assist in a more precise postmortem estimation, since the presence of certain chemical substances may influence the decay process of a cadaver. The increase in the number of drug-related deaths, especially by abuse of heroin and cocaine or deaths connected to accidental or suicidal consumption of poisons or toxic substances, justifies the substantial interest in forensic entomology, especially entomotoxicology.⁹

Insects and other arthropods can be analyzed by standard toxicological procedures, such as radioimmunoassay, gas chromatography, thin-layer chromatography, high performance liquid chromatography, and gas chromatography - mass spectrometry.^{1,8,10} However, forensic toxicology analysis requires simpler and faster methods for the screening of chemical substances that aid in the elucidation of crimes¹⁰ for example, using attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy¹¹ or near infrared spectroscopy.¹²

Electrochemical techniques are powerful and versatile analytical tools that offer high sensitivity, precision, with a large linear dynamic range and low-cost instrumentation. The development of more sensitive electrochemical techniques allows its application in the detection of chemical substances such as drugs, both in its crude form and added to biological samples.¹³ Voltammetric techniques are examples of electrochemical methodologies used in the identification and dosage of pharmaceutical compounds in various forms (tablets, capsules, injections, suspensions) as well as in biological samples.¹³⁻

23

Voltammetric measurements are simple and easily performed.¹³ In this sense, the present report proposes a simple, low cost and high precision electrochemical method to

detect the presence of flunitrazepam in necrophagous larvae using differential pulse voltammetry (DPV). Voltammograms were submitted to multivariate classification analysis – linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) – aided by data dimensionality reduction algorithms: principal component analysis (PCA), successive projections algorithm (SPA) and genetic algorithm (GA). The proposed combination of voltammetry and chemometrics was evaluated in the identification of necrophagous larvae containing flunitrazepam.

Several groups have applied fluorescence spectroscopy to identify spectral characteristics that correspond to the presence of target analytes in complex mixtures. The main focus of these approaches is to identify drug abuse such as flunitrazepam in fly larvae that have fed on a decomposing victim of a crime, the so-called "date rape", where it is possible to compare specimens with and without flunitrazepam through an analysis of their spectral profiles. In this context, the application of multivariate analysis is considered a powerful tool to improve classification or quantification analysis. For example, fluorescence excitation-emission matrix (EEM) has been used in oil samples to monitor the content of polycyclic aromatic compounds,²⁴ developing a method to facilitate discrimination of diesel fuel with rebate tax from oil that is illegally processed by the sorption process²⁵ using EEM fluorescence combined with parallel factor analysis (PARAFAC) to develop a quantitative method for analyzing simulated amphetamine-type illegal drug.²⁶ Satisfactory results were found for the simultaneous determination of methamphetamine and 3,4-ethylenedioxymethamphetamine in the presence of the drug.

This paper reports a method using DPV and fluorescence spectroscopy combined with multivariate analysis to discriminate control and flunitrazepam contamination in larvae

of necrophageous flies. EEM data classification was performed using unfolded LDA,²⁷ QDA²⁷ and support vector machines (SVM).²⁸ These methodologies were compared with three-way EEM data using 2D algorithms, including parallel factor analysis with linear discriminant analysis (PARAFAC-LDA)²⁹ and two-dimensional linear discriminant analysis (2D-LDA).³⁰

Experimental

Animals used in the experiment

The experiments were performed with 32 Wistar rats (*Rattus norvegicus*) having an average weight of 255 ± 30 g (Ethics Committee Approval Protocol 044/2013). The rats with age of 60 ± 5 days used at the beginning of the experiment from the vivarium of the Department of Biophysics and Pharmacology of the Federal University of Rio Grande do Norte were accommodate (4 animals/cage) in nursery cages with free access to water (or ethanol solutions) and food (Purina®, Labina). The rats were divided into four groups with 8 rats each, where each group received different treatment: (1) ethanol – in which ethyl alcohol (Alcoolabor®, Segmenta) was used in increasing concentrations of 2% (3 days), 4% (3 days), 6% (15 days) totaling a period of 21 days;³¹ (2) flunitrazepam – dosage of 2 mg/kg orally;³² (3) conjugated (ethanol plus flunitrazepam) – ethyl alcohol (Alcoolabor®, Segmenta) was used in increasing concentrations of 2% (3 days), 4% (3 days), 6% (15 days) totaling a period of 21 days³¹ and received doses of 2 mg/kg flunitrazepam orally;³² (4) control - received only water *ad libitum*. After one hour of administration of flunitrazepam, the rats were euthanized by guillotine decapitation and each rat was placed

in individualized plastic bags. Animals belonging to the four treatments were transported separately.

In the field, the smell traps containing the rats were distributed on an 8 grid (grid distance of approximately 50 m and at 1.50 m high from the ground) along one of the trails in the 7th Combat Engineering Battalion (The Brazilian Army), an Atlantic forest area located in the city of Natal/Brazil. Each grid contained one rat of each treatment (control, ethanol, flunitrazepam, and conjugate) distant 5 m apart. Traps remained in the field for five consecutive days. On the third and fourth days, 10 immatures were collected (resulting from oviposition, prioritizing the largest larvae, representing the first generations that arrived in the carcass) of each trap. The larvae were placed in a glass vial containing glycerin and sent to the laboratory for identification and chemical analysis.

A set of 3 larvae was added to 2 mL of dimethylsulfoxide (DMSO), then triturated, homogenized and centrifuged. The supernatant solution was stored in a cryotube and then sent for analysis. Two groups were formed for classification, (1) control plus ethanol – larvae collected in carcasses of the control and ethanol groups; (2) flunitrazepam – larvae collected on flunitrazepam and conjugate groups. In total, 40 samples (20 control and 20 containing flunitrazepam) being directed to DPV and fluorescence EEM analysis.

The analyzes were performed with larvae of *Chrysomya megacephala*, *C. albiceps*, *Lucilia sp* and *Cochliomyia macellaria* from the 3rd instar, with specimens of each species in both the control and flunitrazepam groups.

Electrochemical analysis

Differential pulse voltammograms. All DPV measurements were performed by a portable bipotentiostat/galvanostat μ Stat 400 (DropSens, Spain), controlled by DropView software, and carried out in phosphate buffer 0.1 mol L^{-1} at $\text{pH} = 7$, from $+1.0 \text{ V}$ to -1.0 V , scan rate of 15 mV s^{-1} , pulse time of 50 ms and pulse amplitude of 100 mV . The three-electrode system was composed by an Ag/AgCl reference electrode, a graphite lead as the auxiliary electrode, and a carbon nanotubes paste electrode as the working electrode. The following composition was used in the preparation of the working electrode: 55% w:w of multiwall carbon nanotubes (Sigma-Aldrich) and 45% w:w of mineral oil (Specsol). The electrode was assembled by compacting the carbon nanotubes paste at the tip of a 1 mL plastic syringe. The internal electric contact was made by a cleaned copper wire, which also acts as a plunger for paste extrusion. Renewal of the electrode surface was achieved by paste extrusion and smoothing onto a sulfite paper before each new measurement. All voltammograms were obtained in triplicate.

Computational analysis. All voltammetric data was processed using MATLAB[®] R2014 software (Mathworks, Natick, USA) with PLS Toolbox version 7.9.3 (Eigenvector Research, Inc., Manson, USA). Raw voltammograms were pre-processed by cutting between -0.2 V and $+0.9 \text{ V}$, and applying Savitzky–Golay first derivative (window 15 points). Dataset comprising 80 voltammograms was divided into training ($n=52$), validation ($n=14$) and test ($n=14$) sets using the Kennard-Stone (KS) sample selection algorithm.³³ The KS algorithm was applied separately to each class for extract a representative set of objects from a given class by maximizing the minimal Euclidean distance between the selected and the remaining objects. The training samples were used in the modeling procedure, whereas the test set was only used in the final evaluation of the classification.

The optimum number of variables for SPA–LDA, SPA–QDA, GA–LDA, and GA–QDA was based on the average risk G of LDA/QDA misclassification. Such a cost function is calculated in the validation set as:

$$G = \frac{1}{N_V} \sum_{n=1}^{N_V} g_n \quad (1)$$

where N_V is the number of validation samples, and g_n is defined as

$$g_n = \frac{r^2(x_n, m_{I(n)})}{\min_{I(m) \neq I(n)} r^2(x_n, m_{I(m)})} \quad (2)$$

where $I(n)$ is the index of the true class for the n th validation object x_n , $m_{I(n)}$ the samples mean of their true class, and $m_{I(m)}$ the samples mean of the closest wrong class.

Sensitivity (SENS, the confidence that a positive result for a sample of the label class is obtained), specificity (SPEC, the confidence that a negative result for a sample of non-label class is obtained), positive predictive value (PPV, measures the proportion of correctly assigned positive examples), negative predictive value (NPV, measures the proportion of correctly assigned negative examples), Youden's index (YOU, evaluates the classifier's ability to avoid failure), and the likelihood ratios LR+ (the ratio between the probability to predict an example as positive when it is truly positive, and the probability to predict an example as positive when it is actually not positive) and LR- (the ratio between the probabilities to predict an example as negative when it is actually positive, and the probability to predict an example as negative when it is truly negative) were calculated as important quality parameters in test evaluation.

$$\text{SENS}(\%) = \left(\frac{\text{TP}}{\text{TP} + \text{FN}} \right) \times 100 \quad (3)$$

$$SPEC(\%) = \left(\frac{TN}{TN+FP} \right) \times 100 \quad (4)$$

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

$$NPV = \frac{TN}{TN + FN} \quad (6)$$

$$YOU = SENS - (1 - SPEC) \quad (7)$$

$$LR(+) = \frac{SENS}{1 - SPEC} \quad (8)$$

$$LR(-) = \frac{1 - SENS}{SPEC} \quad (9)$$

FN is defined as false negative, FP as false positive, TP as true positive and TN as true negative. SENS is defined as sensibility and SPEC as specificity.

Fluorescence spectroscopy

Emission Excitation Matrix. This procedure was performed for obtaining the individual EEM of control and contaminated with flunitrazepam classes. The sample spectra were acquired with an RF-5301 Shimadzu spectrofluorometer using a 0.5 mm quartz cuvette. The excitation and emission monochromator slit widths were fixed at 1.5 and 3 nm, respectively. The samples were added to a cuvette with a 100 µL micropipette. The cuvette was cleaned by ultra-pure water after each measurement. The temperature was kept at 25 °C throughout the experiments. For larval extract samples, the spectral surfaces of emission/excitation were obtained in the excitation range from 280 to 390 nm (10 nm steps)

and in the emission range from 350 to 900 nm (1 nm steps). This protocol resulted in a data matrix size of 9 x 322 variables for each sample.

Computational analysis. All EEM data was processed using MATLAB® R2014 software with PLS Toolbox version 7.9.3, EEMscat algorithm³⁴ and lab-made routines. The data was firstly pre-processed by removing Rayleigh and Raman scatterings using EEMscat algorithm and then mean centered before chemometric analysis.

The samples were divided into training (n=20), validation (n=10) and test (n=10) sets using the KS sample selection algorithm.³³ The training set was composed of 10 samples of class 1 (control) and 10 samples of class 2 (drugged with flunitrazepam); both validation and test sets had 5 samples of class 1 and 5 samples of class 2. For classification, the data were initially unfolded and the first-order classification algorithms were tested. The unfolded procedure reshaped each EEM matrix with the size of 9 x 322 to a vector of size 1 x 2898. LDA, QDA, and SVM were applied to the data after principal component analysis (PCA) reduction.³⁵ PCA reduced the unfolded data to a few numbers of principal components (PCs), in which the scores on the selected PCs were used as input variables for classification.

In addition, second-order classification algorithms were applied to the original EEM matrices using PARAFAC-LDA and 2D-LDA algorithms. PARAFAC-LDA is based on a PARAFAC³⁶ decomposition of the three-way data of EEM matrices, followed by the application of LDA to the PARAFAC scores matrix. The PARAFAC decomposition is represented by²⁹

$$\underline{\mathbf{X}} = \mathbf{A}(\mathbf{C}|\otimes|\mathbf{B})^T + \underline{\mathbf{E}} \quad (10)$$

in which $\underline{\mathbf{X}}$ represents the three-way EEM matrices; \mathbf{A} is the PARAFAC scores; \mathbf{B} is the PARAFAC loadings representing the excitation direction; \mathbf{C} is the PARAFAC loadings representing the emission direction; $\underline{\mathbf{E}}$ is the residual tensor; and \otimes represents the Khatri-Rao product.

The 2D-LDA algorithm is an adaptation of LDA to three-way data. The classification of a test sample is based on assigning its feature matrix \mathbf{Y}_{test} to the class p^* with the smallest average distance:³⁷

$$\bar{d}(\mathbf{Y}_{test}, C_{p^*}) = \min_{p=1,2,\dots,L} \bar{d}(\mathbf{Y}_{test}, C_p) \quad (11)$$

in which $\bar{d}(\mathbf{Y}_{test}, C_p)$ is the average distance between the test sample and the N_p training samples of class C_p :

$$\bar{d}(\mathbf{Y}_{test}, C_p) = \frac{1}{N_p} \sum_{k \in I_p} d(\mathbf{Y}_{test}, \mathbf{Y}_k) \quad (12)$$

where $d(\mathbf{Y}_{test}, \mathbf{Y}_k)$ is the Euclidian distance between \mathbf{Y}_{test} and the feature vector \mathbf{Y}_k of the samples in the training set. The feature vector \mathbf{Y} of each sample is calculated for a given data matrix \mathbf{X} as:

$$\mathbf{Y} = \mathbf{XB} \quad (13)$$

where \mathbf{B} is a projection matrix obtained by maximizing Fisher's linear projection criterion; thus, maximizing the between-class over the within-class scatter matrices.³⁷

Statistical validation. The statistical validation of the models built was evaluated according to the accuracy, sensitivity and specificity parameters. The accuracy represents

the total number of samples correctly classified considering true and false negatives. This parameter is calculated as follows:³⁸

$$\text{Accuracy}(\%) = \left(\frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \right) \times 100 \quad (14)$$

where TP stands for true positive; TN for true negative; FP for false positive; and FN for false negative.

Results and Discussion

Electrochemical analysis

In total, 80 voltammograms were acquired. The average differential pulse voltammograms for each class in the range of -1.0V to $+1.0\text{V}$ presented similar shape (control, black line; flunitrazepam, red line, Fig. 1a). Baseline correction was performed in the voltammograms by using a B-spline interpolation function after baseline points established by first derivative Savitzky-Golay smoothing filter with a first-order polynomial (3-points windows size). Fig. 1b shows the baseline-corrected voltammograms.

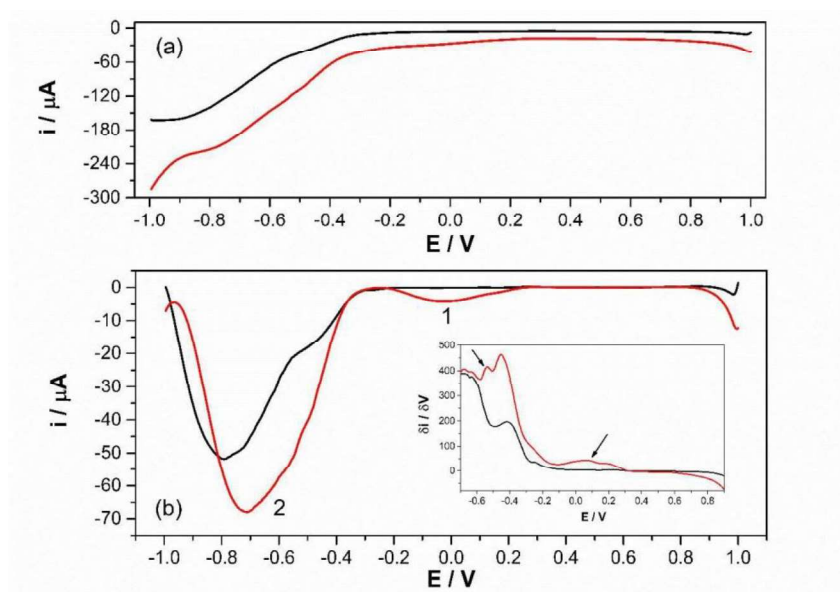
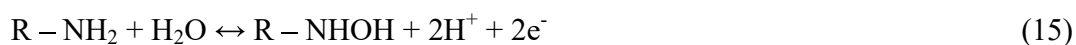


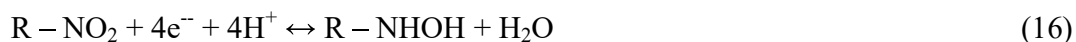
Figure 1. Average DPV data for control (black) and flunitrazepam-contaminated (red) samples (a) before and (b) after baseline correction. Insert: voltammograms transformed by first-derivative. Arrows indicate the main differences in the mean voltammogram for each class.

The voltammogram related to flunitrazepam-contaminated samples shows two important differences compared to the voltammogram of the control samples. First, a broad less intense peak between -0.2 V and $+0.2$ V, which could be attributed to a redox process of 7-aminoflunitrazepam,³⁹ a pharmacologically active metabolite of flunitrazepam. The scan initiates at $+1.0$ V, promoting the oxidation of the amine group to hydroxylamine, which is back reduced to the amine, as described by Equation 15, and giving rise to peak 1 (Fig. 1b).



The second difference is related to peak 2, composed of two redox processes partially overlapped in the voltammogram of control samples, related to the composition of

sample matrix. In the average voltammogram of samples contaminated with flunitrazepam, peak 2 appears with higher current intensity and peak potential, which could be due to the contribution of the reduction of 7-nitro group of non-metabolized flunitrazepam to hydroxylamine³² according to Equation 16.



Insert of Fig. 1b shows the average voltammograms after Savitzky–Golay first derivative. One can see an additional peak between -0.3 V and -0.7 V in the voltammogram of contaminated samples, confirming a third redox process related to the presence of flunitrazepam. It is important to point out that the best results of classification were obtained with voltammograms transformed by first derivative and potential window from -0.2V to $+0.9\text{V}$, i.e., by using information related only to peak 1. Consequently, in postmortem specimens, the drug metabolized over time to 7-aminoflunitrazepam makes it a critical analyte for this study.

PCA was applied as an unsupervised exploratory method to visualize the distribution of the samples in the multidimensional space. Figure 2 presents the PCA scores plot. One can observe that approximately 98% of data variability was explained by two principal components (PC), and there is a substantial dispersion and overlapping of the control and flunitrazepam-contaminated samples. Following, the application of supervised algorithms of classification, more specifically LDA or QDA, was performed to build multivariate classification models. These were adopted systematically to discriminate control vs. contaminated samples based on DPV.

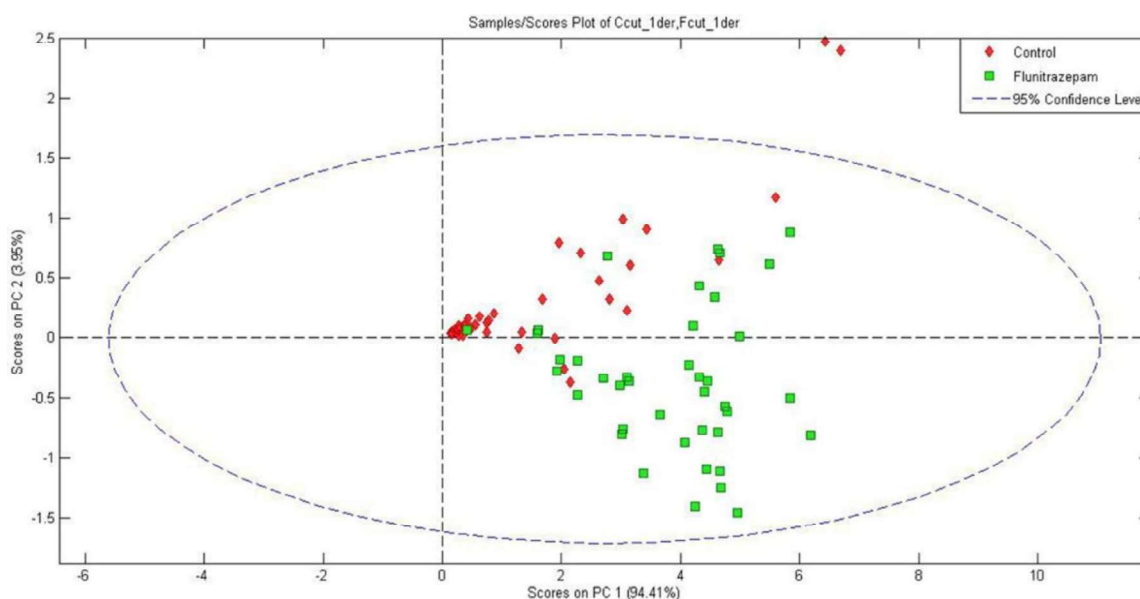


Figure 2. Exploratory analysis with PCA: PC1xPC2 scores plot.

Table 1 presents the figures of merit of the obtained classification models for each class. According to the results shown in Table 1 for control samples, it is possible to conclude that this class presents a more homogeneous composition measured by DPV, always reaching ideal scores by any of the evaluated methods. On the other hand, for flunitrazepam-contaminated class, the figures of merit obtained with QDA models were better than those obtained with the LDA ones. This could be related to the greater heterogeneity of such class due to the more complex process of assimilation and metabolization of flunitrazepam by larvae, which required a quadratic discriminant function to perform a better classification. For example, sensitivity values of QDA models were superior to that obtained with LDA, being 92.9%, 92.9%, and 100% for PCA-QDA, SPA-QDA, and GA-QDA, respectively. Considering all figures of merit of Table 1, GA-QDA model presented the maximum efficiency for classification.

Table 1. Figures of merit of LDA and QDA classification models for each class. Sens: sensitivity; spec: specificity; ppv: positive predictive value; npv: negative predictive value; you: Youden's index; LR(+): positive likelihood ratio; LR(-): negative likelihood ratio.

Algorithm	Control		Flunitrazepam		Algorithm	Control		Flunitrazepam	
PCA-LDA	sens	100	sens	78.6	PCA-QDA	sens	100	sens	92.9
	spec	100	spec	78.6		spec	100	spec	100
	ppv	100	ppv	78.6		ppv	100	ppv	100
	npv	100	npv	78.6		npv	100	npv	93.3
	you	100	you	57.1		you	100	you	92.9
	LR(+)	0.0	LR(+)	0.0		LR(+)	0.0	LR(+)	0.0
	LR(-)	0.0	LR(-)	0.3		LR(-)	0.0	LR(-)	0.1
SPA-LDA	sens	100	sens	85.7	SPA-QDA	sens	100	sens	92.9
	spec	100	spec	85.7		spec	100	spec	100
	ppv	100	ppv	85.7		ppv	100	ppv	100
	npv	100	npv	85.7		npv	100	npv	93.3
	you	100	you	71.4		you	100	you	92.9
	LR(+)	0.0	LR(+)	0.0		LR(+)	0.0	LR(+)	0.0
	LR(-)	0.0	LR(-)	0.2		LR(-)	0.0	LR(-)	0.1
GA-LDA	sens	100	sens	78.6	GA-QDA	sens	100	sens	100
	spec	100	spec	100		spec	100	spec	100
	ppv	100	ppv	100		ppv	100	ppv	100
	npv	100	npv	82.4		npv	100	npv	100
	you	100	you	78.6		you	100	you	100
	LR(+)	0.0	LR(+)	0.0		LR(+)	0.0	LR(+)	0.0
	LR(-)	0.0	LR(-)	0.2		LR(-)	0.0	LR(-)	0.0

The excellent discrimination power obtained for GA-QDA model is presented in Fig. 3. GA-QDA model employed 13 variables (Fig. 4). Four selected potentials seem to be of particular interest. Variables -0.02 V, $+0.14$ V, $+0.28$ V, and $+0.30$ V are located at peak 1, and model the influence of the presence of the metabolite 7-aminoflunitrazepam in differentiating each class with success. The other nine variables are located at the baseline and model the change on background current regarding the presence of flunitrazepam and its metabolite.

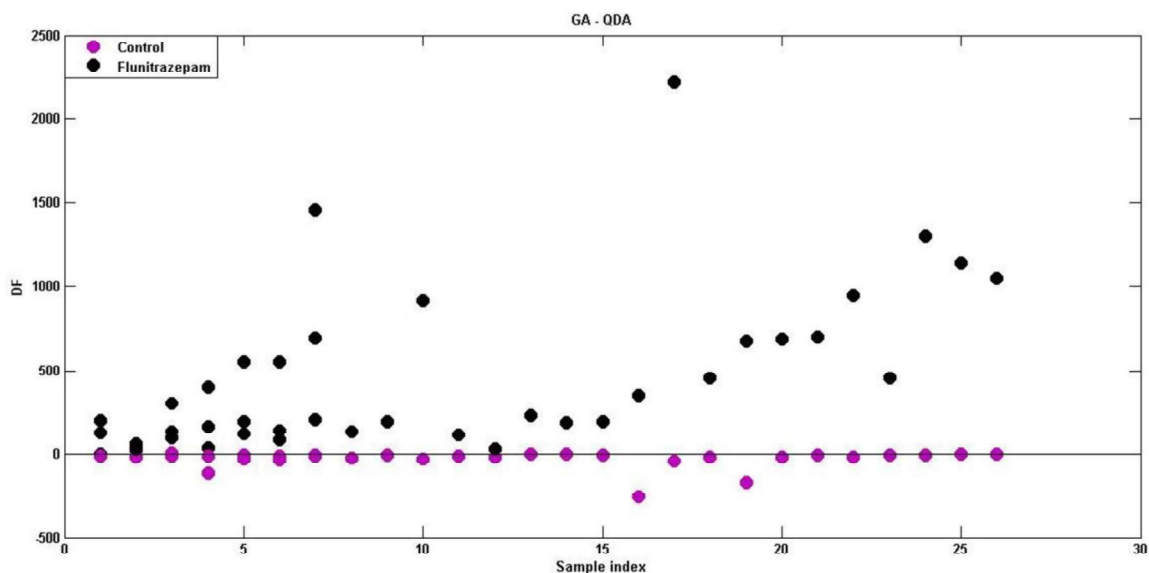


Figure 3. The Discriminant function calculated by GA-QDA model for control (blue) *versus* flunitrazepam-contaminated (black) samples.

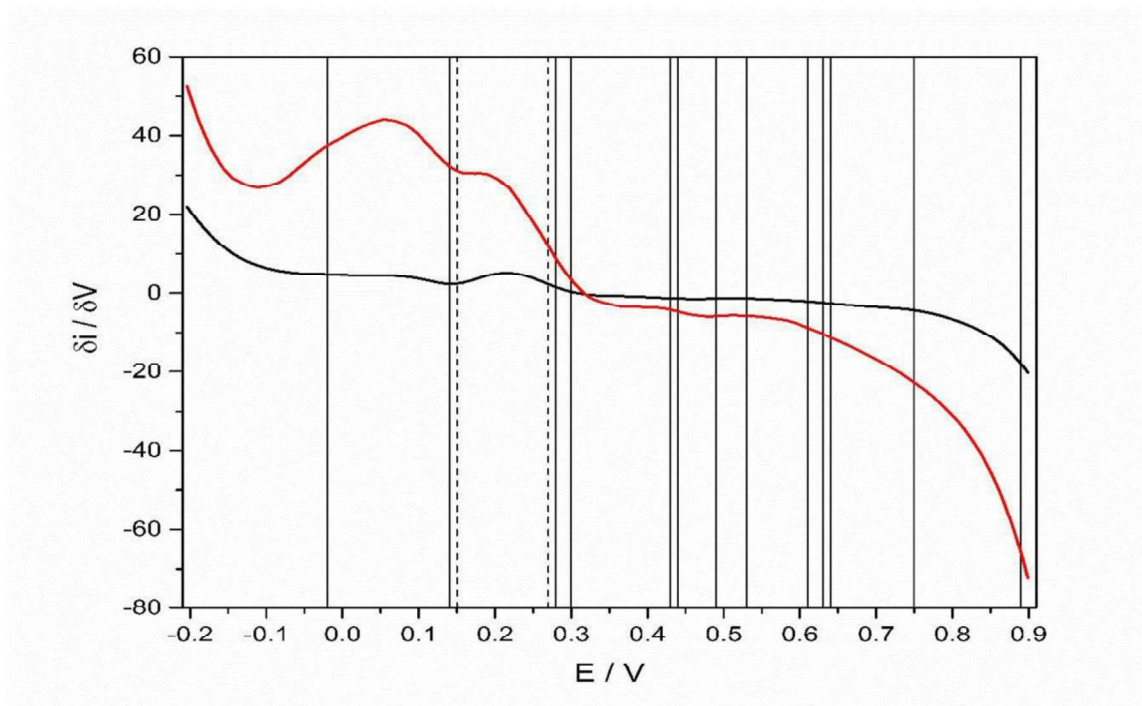


Figure 4. Average first-derivative voltammograms of each class (black: control; red: flunitrazepam) and the variables selected by GA-QDA (solid vertical lines) and SPA-QDA (dashed vertical lines).

SPA-QDA model employed two variables (Fig. 4). Despite the lower classification efficiency offered by the SPA-QDA model, it can be considered that its results were satisfactory given the lower number of variables selected, which implies greater robustness when compared to GA-QDA. Examination of the selected potentials showed that the model used information associated to the presence of the drug detected by DPV in the contaminated samples.

Fluorescence analysis

Average EEM data for class 1 (control) and class 2 (larvae with flunitrazepam) are shown in Figure 5. As can be seen, a visual inspection cannot reveal significant differences between the two classes, since both profiles seem very similar.

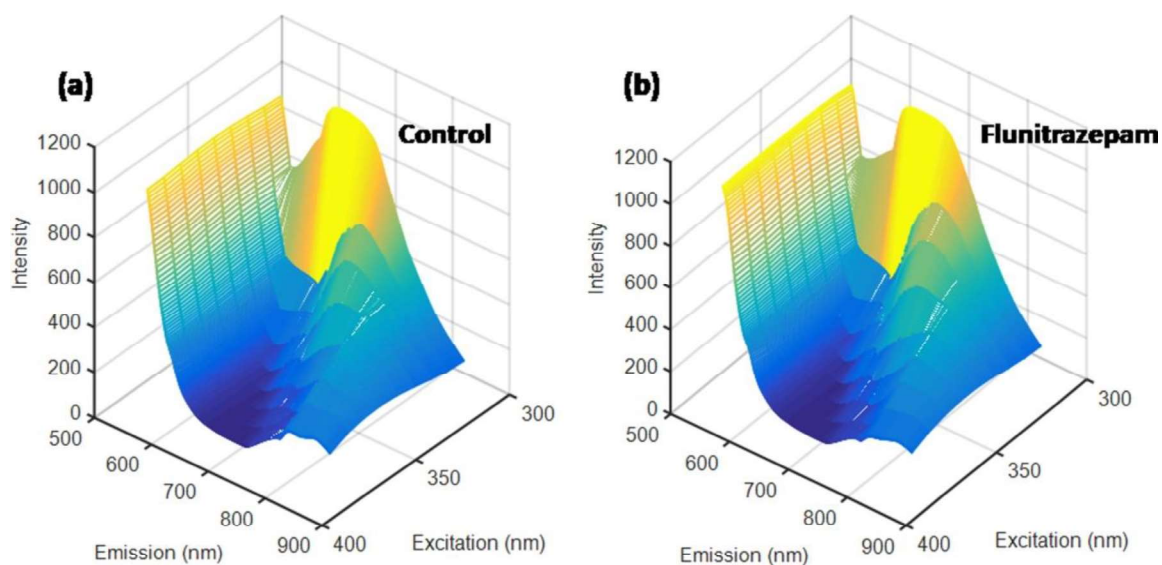


Figure 5. Average EEM data for control (a) and flunitrazepam-contaminated (b) samples.

A closer look into the excitation and emission profiles for each EEM can be seen in Figure 6. As demonstrated by Figure 6, both excitation and emission profiles for control and flunitrazepam contaminated samples are much superposed among the samples, where no clear separation can be observed.

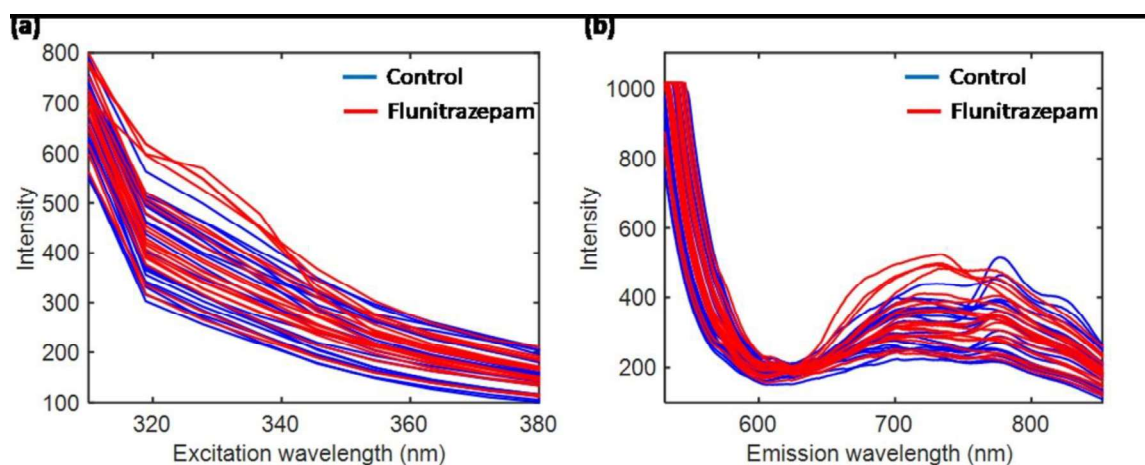


Figure 6. Excitation (a) and emission (b) profiles for control (blue) and flunitrazepam-contaminated (red) samples.

For an initial assessment, each EEM data matrix was unfolded into a vector and first-order classification algorithms were tested. LDA, QDA and SVM were applied to the PCA scores on the first 10 PCs (99.36% of cumulative variance). The results are shown in Table 2.

Table 2. Figures of merit calculated using different classification methods applied to the PCA scores of unfolded EEM data.

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
LDA	60.0	60.0	60.0
QDA	50.0	20.0	80.0
SVM	50.0	40.0	60.0

Table 2 shows that first-order classification algorithms did not successfully discriminate the two classes. The best accuracy and sensitivity was found with LDA (60%) and the best specificity with QDA (80%). However, overall these values are unsatisfactory for entomotoxicology applications, since the gold method using HPLC reaches values above 90% sensitivity and specificity. New methodologies such as 2D-LDA are strongly able to achieve the efficiency of gold methodologies and require lower cost, less sample preparation and analysis time, and keeps the analytical accuracy at acceptable levels for expert requirements.

Second-order classification algorithms were applied by means of PARAFAC-LDA and 2D-LDA. PARAFAC-LDA was built with 5 components, and 2D-LDA was built with 2 factors. Their classification results are shown in Table 3.

Table 3. Figures of merit calculated using PARAFAC-LDA and 2D-LDA applied to the three-way EEM data.

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
PARAFAC-LDA	50.0	40.0	60.0
2D-LDA	100	100	100

The overall classification performance of PARAFAC-LDA was very similar to that obtained using first-order algorithms (Table 2), in which poor discrimination between control and flunitrazepam-contaminated samples were observed. On the other hand, 2D-LDA presented a perfect performance with classification rates equal to 100% for all figures of merit. The 2D-LDA DF plot is shown in Figure 3, where a perfect separation of the two classes is clearly observed.

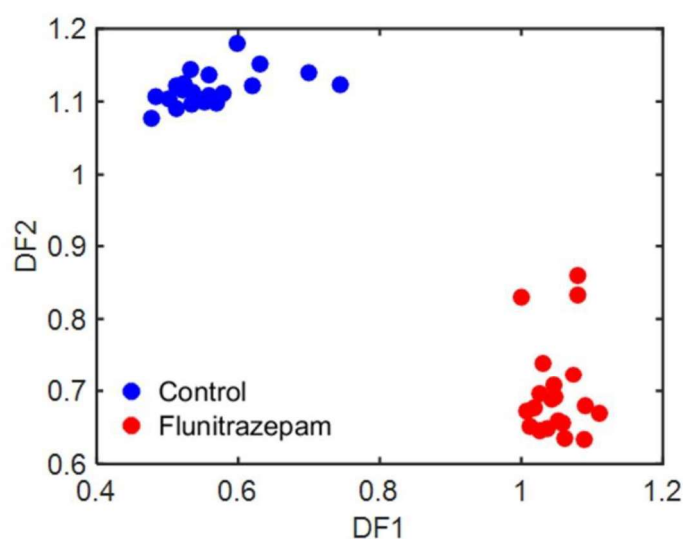


Figure 7. Discriminant function (DF) plot for factor 1 *versus* factor 2.

2D-LDA has been reported as being superior to other classification methods such as PARAFAC in many chemical applications.³⁷ Second-order classification algorithms seem

superior to their unfolded versions due to the loss of spatial distribution information during unfolding procedure.³⁸ Therefore, the use of two-dimensional classification algorithms is the best option to analyze EEM data.

Conclusions

The results of this study show that differential pulse voltammetry combined with multivariate classification may be an alternative tool for the detection of flunitrazepam in necrophagous larvae. We report a fast, clean and low-cost method, which involves minimum samples preparation to classify the specimens. In this report, the resulting GA-QDA model successfully detected flunitrazepam providing 100% accuracy and specificity. This method was completely validated, showing potential for use as an official methodology for entomological methods. EEM fluorescence spectroscopy with 2D-LDA was effective to differentiate control and larval extract drugged with flunitrazepam for an entomotoxicology application. Using this methodology, 100% accuracy, sensitivity and specificity were found for discriminating the two classes. It represents a fast, precise and low-cost method for forensic analysis of larval extract, having a minimum sample preparation and with a small amount of sample required for analysis. This study is restricted to a local level application due to the number of samples used in the experiment and the environmental conditions submitted to the insects. Future studies are needed for the purpose of increasing sample space aiming at a global model, and may enable the analysis of new drugs for forensic interest.

Acknowledgments

Leomir A. S. Lima would like to acknowledge the financial support from PPGQ/UFRN/CAPES for his fellowship. Camilo L. M. Morais would like to

thank CAPES-Brazil (Doutorado Pleno no Exterior, grant 88881.128982/2016-01) for financial support. Kássio M. G. Lima would like to acknowledge the CNPq grant (305962/2014-0) for financial support. Sherlan G. Lemos would like to acknowledge the CNPq grant (301662/2015-0) for financial support.

References

1. Amendt, J.; Campobasso, C. P.; Gaudry, E.; Reiter, C.; LeBlanc, H. N.; Hall, M. J. R.; *J. Leg. Med.* **2007**, *121*, 90.
2. Catts, E. P.; *J. Agric. Entomol.* **1992**, *9*, 245.
3. Lord, W. D.; Rodriguez, W. C.; *Prosecutor* **1989**, *22*, 41.
4. Benecke, M.; *Forensic Sci. Int.* **2001**, *120*, 2.
5. De Santana, C. S.; Siquieroli, D.; Boas, V.; *Ceciliana* **2012**, *4*, 31.
6. Gosselin, M.; Wille, S. M. R.; Fernandez, M. M. R.; Di Fazio, V.; Samyn, N.; De Boeck, G.; Bourel, B.; *Forensic Sci. Int.* **2011**, *208*, 1.
7. Wallace, D. R.; *Tox. Forensic Med.* **2017**, SE(1): Se1-Se4. doi: 10.17140/TFMOJ-SE-1-e001.
8. Introna, F.; Campobasso, C. P.; Goff, M. L.; *Forensic Sci. Int.* **2001**, *120*, 42.
9. Goff, M.; Lord, W.; *J. Forensic Med. Path.* **1994**, *15*, 51.
10. Smith, M. L.; Vorce, S. P.; Holler, J. M.; Shimomura, E.; Magluilo, J.; Jacobs A. J.; Huestis, M. A.; *J. Anal. Tox.* **2007**, *31*, 237.
11. Oliveira, J. S.; Baia, T. C.; Gama, R. A.; Lima, K. M. G.; *Microchem. J.* **2014**, *115*, 39.
12. de Lima, L. A. S.; Baia, T. C.; Gama, R. A.; Lima, K. M. G.; *NIR news* **2014**, *25*, 23.
13. Huitle, C. A. M.; Fernandes, N. S.; Lopes, M. C.; Quiroz, M. A.; *Port. electrochi. acta.* **2010**, *28*, 39.
14. Bagalkoti, J. T.; Pattar, V. P.; Nandibewoor, S. T.; *J. Electroche. Sci.Eng.* **2017**, *7*, 77.
15. Batchelor-McAuley, C.; Katelhon, E.; Barnes, E. O.; Compton, R. G.; Laborda, E.; Molina, A.; *Chem. Open.* **2015**, *4*, 224.
16. Branica, G.; Metiko, M.; Omanovi, D.; *Croat. Chem. Acta* **2006**, *79*, 77.

17. Cuculi, V.; Pi, I.; Branica, M.; *J. Electroanal. Chemi.* **2004**, 583, 1.
18. Farghaly, O. A. E.; *J. pharm. biomed. anal.* **2000**, 23, 783.
19. Farghaly, O. A.; Hameed, R. S. A.; *J. Electroanal. chemi.* **2014**, 25, 37.
20. Frenke, J. P.; De Zeeuw, R. A.; *J. Anal. Tox.* **1976**, 1, 291.
21. Hoang, V. D.; Huyen, D. T.; Phuc, P. H.; *J. anal. methods chemi.* **2013**, 2013.
22. Yardim, Y.; Levent, A.; Zühre, S.; *Talanta* **2011**, 85, 441.
23. Zayed, S. I. M.; Issa, Y. M.; *Bioelectrochemistry* **2009**, 75, 9.
24. Christensen, J. H.; Hansen, A. B.; Mortensen, J.; Andersen, O.; *Anal. Chemi.* **2005**, 77, 2210.
25. Orzel, J.; Daszykowski, M.; Grabowski, I.; Zaleszczyk, G.; Sznajder, M.; *Fuel* **2014**, 117, 224.
26. Xu, B.; Ye, Y.; Liao, L.; *Res. Forensic Sci.* **2017**, 1.
27. Wu, W.; Mallet, Y.; Walczak B.; Penninckx, W.; Massart, D. L.; Heuerding, S.; Erni, F.; *Anal. Chimi. Acta* **1996**, 329, 257.
28. Cortes, C.; Vapnik, V.; *Mach. Learn.* **1995**, 20, 273.
29. Costa, F. S. L.; Silva, P. P.; Morais, C. L. M.; Theodoro, R. C.; Arantes, T. D.; Lima, K. M. G.; *Anal. Methods* **2017**, 9, 3968.
30. Li, M.; Yuan, B.; *Pattern Recog. Lett.* **2005**, 26, 527. doi:10.1016/j.patrec.2004.09.007.
31. Tirapelli, C. R.; Fukada, S. Y.; Yogi, A.; Chignalia, A. Z.; Tostes, R. C.; Bonaventura, D.; Lanchote, V. L.; Cunha, Q.; De Oliveira, A. M.; *Braz. J. Pharm.* **2008**, 153(3), 468.
32. Shinomiya, K.; Shigemoto, Y.; Omichi, J.; Utsu, Y.; Mio, M.; Kamei, C.; *Psychopharmacol.* **2004**, 173(1), 203.
33. Kennard, R. W.; Stone, L. A.; *Technometrics* 1969, 11, 137.
34. Bahram, M.; Bro, R.; Stedmon, C.; Afkhami, A.; *J. Chemom.* **2006**, 20, 99.
35. Bro, R.; Smilde, A. K.; *Anal. Methods* **2014**, 6, 2812.
36. Bro, R.; PARAFAC. *Chemom. Intell. Lab. Syst.* **1997**, 38, 149.

37. Da Silva, A. C.; Soares, S. F. C.; Insausti, M.; Galvão, R. K. H.; Band, B. S. F.; De Araújo. M. C. U.; *Anal. Chim. Acta* **2016**, 938, 53.
38. Morais, C. L. M.; Lima, K. M. G.; *Chemom. Intell. Lab. Syst.* **2017**, 170, 1.
39. Garcia-Gutierrez, E.; Lledo-Fernandez, C.; *Chemosensors* **2013**, 1, 68; doi:10.3390/chemosensors1030068.

Capítulo 5 - Conclusão e perspectivas

Nesta tese, foi proposta a aplicação de técnicas espectroscópicas, eletroanalítica e multivariada para avaliar o desempenho de tratamento para reparo ósseo, classificação de idosos com sarcopenia e identificação de larvas de moscas contendo o fármaco flunitrazepam.

A espectroscopia NIR com algoritmos de classificação multivariada PCA-LDA, SPA-LDA e GA-LDA foram aplicadas em defeito crítico provocado em crânio de ratos para avaliar o desempenho dos tratamentos com colágeno e laser de baixa intensidade frente a um conjunto de crânio controle (Capítulo 2). O melhor desempenho foi obtido com o GA-LDA conseguindo classificar as classes e principalmente a classe controle das demais. Essa investigação indica que a espectroscopia NIR tem bom desempenho na discriminação de diferentes estímulos em crânio de ratos e pode oferecer potencial de diagnóstico não invasivo in vivo de defeito de osso craniano em um futuro próximo.

A espectroscopia ATR-FTIR foi aplicada juntamente às técnicas multivariadas PCA-LDA, SPA-LDA e GA-LDA para a classificação de plasmas sanguíneos de idosos saudáveis e com sarcopenia (Capítulo 3). Os resultados mostraram que a metodologia proposta tem potencial para classificar as amostras com valores de sensibilidade e especificidade superior a 60%. A metodologia proposta é rápida, não invasiva e possível de ser inserida em unidades de saúde básica. Também, elimina algumas etapas que podem ser influenciadas pela subjetividade do método de referências clínica, como as equações de Visvanathan e a equação Lee. Ainda, foi possível selecionar variáveis com os algoritmos SPA e GA que indicam a região de proteína como sendo o principal grupo bioquímico que difere as classes estudadas, indicando um caminho para a investigação de possíveis biomarcadores para sarcopenia.

A fluorescência molecular e a voltametria de pulso diferencial foram aplicadas para identificar flunitrazepam em larva de mosca com o propósito de desenvolvimento de uma metodologia simples, rápida e de baixo custo (Capítulo 4). Por fluorescência foi avaliado o desempenho do algoritmo 2D-LDA frente aos tradicionais algoritmos utilizados para processar dados de segunda. O PARAFAC-LDA obteve taxa de acerto abaixo de 60% enquanto o 2D-LDA alcançou a taxa de acerto de 100%, se mostrando mais eficiente para esse conjunto de dados.

Aplicando voltametria de pulso diferencial foi obtida a melhor classificação, utilizando a análise discriminante quadrático. O modelo GA-QDA resultante classificou com sucesso o flunitrazepam com base em 13 valores de potenciais e alcançou 100% de sensibilidade e especificidade. Os estudos descritos nos capítulos 4 dessa tese são grandes avanços na área da entomotoxicologia forense proporcionando métodos analíticos alternativos e com elevada taxa de acerto. Depois de validados, esses estudos podem ser inseridos na rotina da perícia de institutos de polícia científica podendo ser considerado um retorno à sociedade.

Os estudos propostos mostraram através de seus resultados a possibilidade de introdução da bioespectroscopia e métodos eletroanalíticos em conjunto com algoritmos de classificação multivariada como ferramenta para o estudo de sistemas biológicos complexos. Entretanto, é necessário expandir esses estudos, principalmente no sentido de ampliar o número de amostras analisadas e contemplar uma maior variabilidade dos dados, durante a etapa de validação. Ainda, as abordagens relatadas nesta tese podem ser extrapoladas para outros tipos de problemáticas envolvendo sistemas biológicos como: investigação de outros tipos de doenças, identificação de gênero e espécies de insetos, identificação de carga parasitária em hospedeiro e identificação de substâncias tóxicas em mamíferos e insetos.

Apêndice A

Near infrared spectroscopy as an emerging tool for forensic entomotoxicology.

Leomir A. S. de Lima

Tainá C. Baia

Renata A. Gama

Luiz Henrique da Silva Gasparotto

Kássio M.G. Lima.

NIRnews, 2014, vol. 25, 8, 5 – 7.

Contribuição:

- Realizei o processamento dos dados e construção dos modelos multivariados;
- Ajudei na escrita da primeira versão do manuscrito.

Leomir A. S. de Lima

Prof. Kássio M. G. Lima

Near infrared spectroscopy as an emerging tool for forensic entomotoxicology

Leomir Aires Silva de Lima,^a Tainá C. Baia,^b Renata A. Gama,^b Luiz Henrique da Silva Gasparotto^a and Kássio M.G. Lima^{a*}

^aUniversidade Federal do Rio Grande do Norte, Instituto de Química, Programa de Pós-Graduação em Química, Grupo de Pesquisa em Química Biológica e Quimiometria, CEP 59072-970 – Natal, RN, Brazil. E-mail: kassiolima@gmail.com.br

^bUniversidade Federal do Rio Grande do Norte, Departamento de Microbiologia e Parasitologia, CEP 59072-970 – Natal, RN, Brazil

Forensic entomotoxicology

Forensic entomotoxicology studies the usefulness of insects as alternative toxicological samples. Investigations of the use of carrion-feeding arthropods as alternative toxicological specimens, and of the impact the tissue toxins and drugs have on the development of immature insects feeding on these substances, currently comprise the major avenues of exploration in the emerging field of entomotoxicology.¹ The major interest of entomotoxicology is the determination of drug abuse just before death, especially in skeletonised remains in which no tissue or fluids remain. Use of insects as an alternative matrix for drug detection is well documented and recommended when conventional matrices such as blood, urine or internal organs are no longer available. Insects are of interest as they are present in high quantities and their remains are present for a long time, even when toxicological samples are no longer available.

NIR as an emerging tool

Near infrared (NIR) spectroscopy can be utilised to determine the insect metabolic fingerprint (lipids, proteins, cellular processes), emerging as an interesting alternative for a rapid and cost-effective identification of living specimens. Near infrared spectroscopy is emerging as the tool of choice for non-destructive analysis and the detection of different compounds in biological systems. NIR should be able to differentiate between insect species based on their absorbance characteristics, because the cuticle of each insect species may have a unique chemical composition² (Figure 1). The absorption is affected by the internal and external biochemical composition of the organism, and different organisms can have unique absorption spectra. NIR, which quantitatively measures organic compounds, e.g., O–H, N–H and C–H functional groups (stretching and bending modes), may reveal unique absorption of NIR energy between

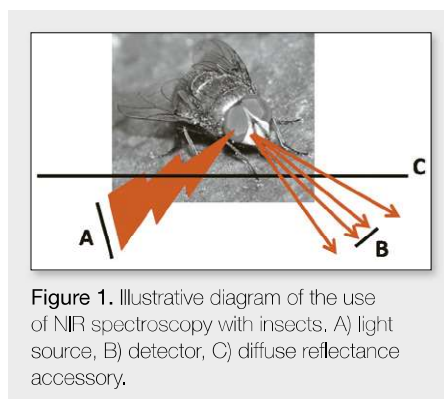


Figure 1. Illustrative diagram of the use of NIR spectroscopy with insects. A) light source, B) detector, C) diffuse reflectance accessory.

species and this uniqueness in the cuticle composition may affect the NIR spectra, and thus can be used in classification models.

Data handling approaches

For the analysis of the insect metabolic fingerprint with NIR, complex datasets are generated, giving rise to the challenges of extracting meaningful underlying variance between and within variables. Similarly, the interrogation of these metabolic insect samples generates complex spectral datasets and necessitates the application of suitable data-handling tools in order to extract important discriminating information. Principal component analysis (PCA) is usually used for initial data reduction during which it simplifies a dataset via linear transformations by choosing a new coordinate system in which the first principal component (PC) describes the greatest variance within the dataset. Alternatively, hierarchical cluster analysis (HCA) may be applied to analyse groups in a dataset on the basis of their spectral similarities, while linear discriminant analysis (LDA) may be used to reduce confounding factors of within-category heterogeneity whilst maximising between-category discriminating biomarkers or to classify unknown samples into predetermined groups. A well-developed approach to identify biomarkers or wavelengths is the successive projection algorithm (SPA) or

genetic algorithm (GA) in conjunction with LDA. Basically, SPA-LDA and GA-LDA employ a cost function associated with the average risk of misclassification in a validation set and can also reduce the generalisation problems often associated with collinearity and avoid over-fitting.³

Objectives

We evaluate the potential of a novel, non-destructive method for identification of flunitrazepam in *Lucilia eximia* (Diptera: Calliphoridae) larvae. For this, the present paper sets out to determine biochemical intra-individual differences or “fingerprint” features between a series of larvae that could be identified using NIR spectroscopy with subsequent variable selection methods; calibrators were prepared by spiking the drug-free larvae aliquots with flunitrazepam standard (0pgg⁻¹ and 2mgkg⁻¹) and solvent (alcohol). We employed SPA and GA to select an appropriate subset of wavelengths for LDA. This approach can lead to more selective and specific biochemical detection for forensically relevant insects by vibrational spectroscopy.

Methods

Animals used in the experiment

Thirty-two Wistar mice (*Rattus norvegicus*) with an average weight of 300g were divided in four groups of eight and supplemented in the following way: group 1—ethanol; group 2—flunitrazepam at the dose of 2mgkg⁻¹; group 3—ethanol and flunitrazepam; group 4—water. One hour after the supplementation the mice were sacrificed, individually housed in fly traps and distributed in eight points along a track of a nearby forest. On the third and fourth days 10 larvae were collected from each mouse and analysed with NIR spectroscopy.

Calibrators and quality control samples

The average weight of one larva was estimated at 80mg. Distilled water was added

to a pool of 50 drug-free larvae to give a final volume of 25 mL. A series of larvae calibrators were prepared by spiking the drug-free larvae aliquots with flunitrazepan standard (0pgg^{-1} and 2mgkg^{-1}) and solvent (alcohol). Therefore, we have four categories of larvae: alcohol, alcohol + drug (2mgkg^{-1}), control (water) and drug (2mgkg^{-1}).

NIR spectroscopy

Spectra ($n=237$; 10 replicate scans from each of 24 specimens) were collected from larvae using a miniature scanning Fourier transform spectrometer from ARCspectro ANIR (900–1700 nm; Neuchâtel, Switzerland), by placing larvae individually on their backs on a 6-cm diameter Spectralon plate. The 3-mm diameter bifurcated fibre-optic probe was positioned about 2 mm above the specimen, focusing on the head. The transfectance probe was washed with ethanol (70% v/v) and dried using tissue paper after each sample. Spectral measurements were done in an acclimatised room under a controlled temperature (22°C) and 60% relative air humidity; samples were allowed to equilibrate to this temperature before analysis.

Data analysis

The data import, pre-treatment and construction of chemometric classification models (PCA-LDA, SPA-LDA and GA-LDA) were implemented in MATLAB R2014a software (Mathworks Inc, Natick, MA, USA). NIR spectra were pre-processed by cutting between 900 nm and 1700 nm, Savitzky–Golay smoothing (with a window of seven points) and calculating the second derivative of the Savitzky–Golay polynomial (with a window of seven points). For PCA-LDA, SPA-LDA and GA-LDA models, the samples were divided into training (70%), validation (15%) and prediction sets (15%) by applying the classic Kennard–Stone (KS) uniform sampling algorithm to the NIR spectra. Sample numbers in each set are presented in Table 1. In the GA-LDA model, the mutation and reproduction probabilities were kept constant at 10% and 80%, respectively. The initial population was 120 individuals with 60 generations. The best solution resulting from the three realisations of the GA was retained.

Classification modeling

As can be seen in Figure 2A, mean NIR spectra acquired from four classes (alcohol, A1; alcohol + drug, A2; drug, A3;

Table 1. Number of training, validation and prediction specimens (or spectra) in each category.

Category	Set training	Validation	Prediction
Alcohol	40	10	10
Alcohol + drug (2mgkg^{-1})	40	10	10
Control (water)	39	10	10
Drug (2mgkg^{-1})	38	10	10

control) present a consistent baseline offset and bias. Although these are quite common features in NIR spectra acquired by diffuse reflectance techniques, some pre-treatments need to be performed. Among the pre-processing techniques tested, the one showing better separation of the classes employing PCA, SPA-LDA and GA-LDA was the combination of Savitzky–Golay second derivative (seven points window, second degree polynomial) and Savitzky–Golay smoothing (seven points window).

After processing, the range of 900–1700 nm was submitted to PCA analysis. The PCA model was built from the calibration set using 10 PCs and cumulatively explaining 95% of the variance in the data after applying the pre-processing (Savitzky–Golay smoothing and second derivative). As shown in Figure 2B, the classes overlap. This weak separation was obtained without using the information that the samples belong to four different groups (alcohol, alcohol + drug, control and drug) and is indicative of features of spectral differences.

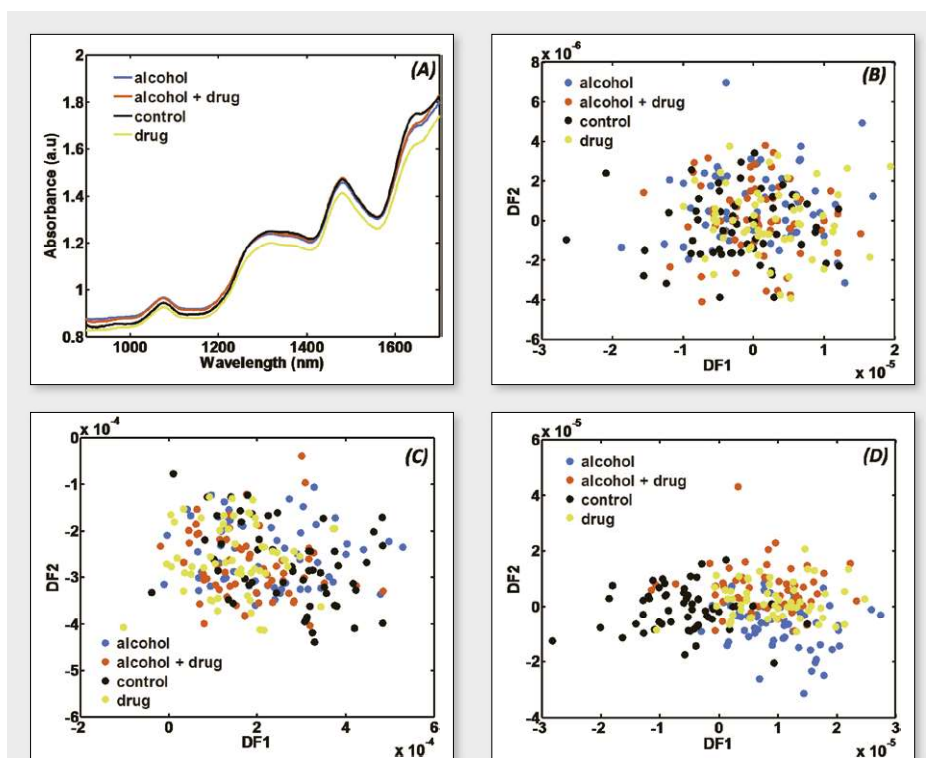


Figure 2. (A) Comparison of biochemical-larvae fingerprint regions for NIR spectroscopy. The panel shows mean NIR spectra obtained from all larvae segregated into alcohol vs alcohol + drug vs control vs drug. The spectra from larva with alcohol, alcohol + drug, control and drug are shown in blue, red, black and yellow, respectively. The remaining panels show the application of principal component analysis (PCA)–linear discriminant analysis (LDA) or variable selection techniques [successive projection algorithm (SPA) and genetic algorithm (GA)] to the segregation of retrospectively categorised larva specimens. (B) PCA-LDA results: DF1 x samples calculated by PCA-LDA model from alcohol (blue), alcohol + drug (red), control (black) and drug (yellow). (C) SPA-LDA results: DF1 x samples calculated using the five selected wavenumbers by SPA-LDA model from alcohol (blue), alcohol + drug (red), control (black) and drug (yellow). (D) GA-LDA results: DF1 x samples calculated using the 17 selected wavenumbers by GA-LDA model from alcohol (blue), alcohol + drug (red), control (black) and drug (yellow).

SPA and GA were applied to larvae data set with four categories: alcohol, alcohol + drug, control and drug.

SPA resulted in the selection of five variables, namely 900, 1317, 1482, 1631 and 1700nm. Using solely these five selected wavelengths, the Fisher scores for all the larvae of the data set were calculated, as shown in Figure 2C. Using only these wavelengths, Fisher scores were obtained and this improved segregation between classes (Figure 2C) when compared with PCA-LDA. Performing LDA on the GA selected variable 17 wavelengths dataset, namely 917, 1007, 1039, 1066, 1161, 1171, 1173, 1181, 1200, 1230, 1352, 1404, 1475, 1491, 1493, 1617 and 1625nm. Using these 17 selected wavelengths, the Fisher scores were obtained for all the specimens in the dataset (Figure 2D) whose cost function minimum point was achieved with 17 wavelengths. As shown, there was a good separation for each category, especially for the control category. However, there is a slight overlap between alcohol and alcohol + drug categories. Examination of the selected wavelengths following GA-LDA indicated that the main biochemical alterations induced by *Lucilia eximia* (Diptera: Calliphoridae) were on cuticular lipids (917, 1007 and 1039nm). Several selected

wavelengths appear to be of particular interest, namely, the variables at 1617nm and 1625nm, representing the first overtone C–H stretching in the flunitrazepam band (fingerprint region). These findings suggest that NIR is a very promising technique for the non-destructive identification of flunitrazepam in *Lucilia eximia* (Diptera: Calliphoridae).

Conclusions

As was demonstrated, NIR spectroscopy facilitates the derivation of an integrated biochemical fingerprint of an insect based on the NIR absorbing properties of the constituent chemical vibrations. The potential of NIR spectroscopy for identification of flunitrazepam in *Lucilia eximia* (Diptera: Calliphoridae) larvae has been examined in this work. The study demonstrates that different categories of larvae can be clearly segregated using NIR spectroscopy with subsequent PCA, SPA-LDA and GA-LDA algorithms.

Acknowledgements

K.M.G. Lima acknowledges the CNPq/Capes project (Grant 070/2012) and FAPERN (Grant 005/2012) for financial support. Post-Graduate Program in Chemistry (PPGQ) at UFRN. T.C. Baia thanks the

Parasitology Biology Post-graduation Program at UFRN. The authors would like to acknowledge the financial support from the CAPES for a fellowship for L.A.S. Lima, the Post-Graduate Program in Chemistry (PPGQ) of UFRN and Parasitology Laboratories.

References

1. M. Gosselin, S.M.R. Wille, M.D.M.R. Fernandez, V. Di Fazio, N. Samyn, G. De Boeck and B. Bourel, "Entomotoxicology, experimental set-up and interpretation for forensic toxicologists", *Forensic Sci. Int.* **208**, 1–9 (2011). doi: <http://dx.doi.org/10.1016/j.forsciint.2010.12.015>
2. J.S. Oliveira, T.C. Baia, R.A. Gama and K.M.G. Lima, "Development of a novel non-destructive method based on spectral fingerprint for determination of abused drug in insects: An alternative entomotoxicology approach", *Microchem. J.* **115**, 39–46 (2014). doi: <http://dx.doi.org/10.1016/j.microc.2014.02.009>
3. A.S. Marques, M.C.N. de Melo, T.A. Cidral and K.M.G. de Lima, "Feature selection strategies for identification of *Staphylococcus aureus* recovered in blood cultures using FT-IR spectroscopy successive projections algorithm for variable selection: A case study", *J. Microbiol. Methods.* **98**, 26–30 (2013). doi: <http://dx.doi.org/10.1016/j.mimet.2013.12.015>

Apêndice B

FTIR microspectroscopy coupled with variable selection methods for the identification of flunitrazepam in necrophagous flies.

Tainá C. Baia

Leomir A. S. de Lima

Renata A. Gama

Kássio M.G. Lima.

Analytical Methods, 2016, 8, 968–972.

Contribuição:

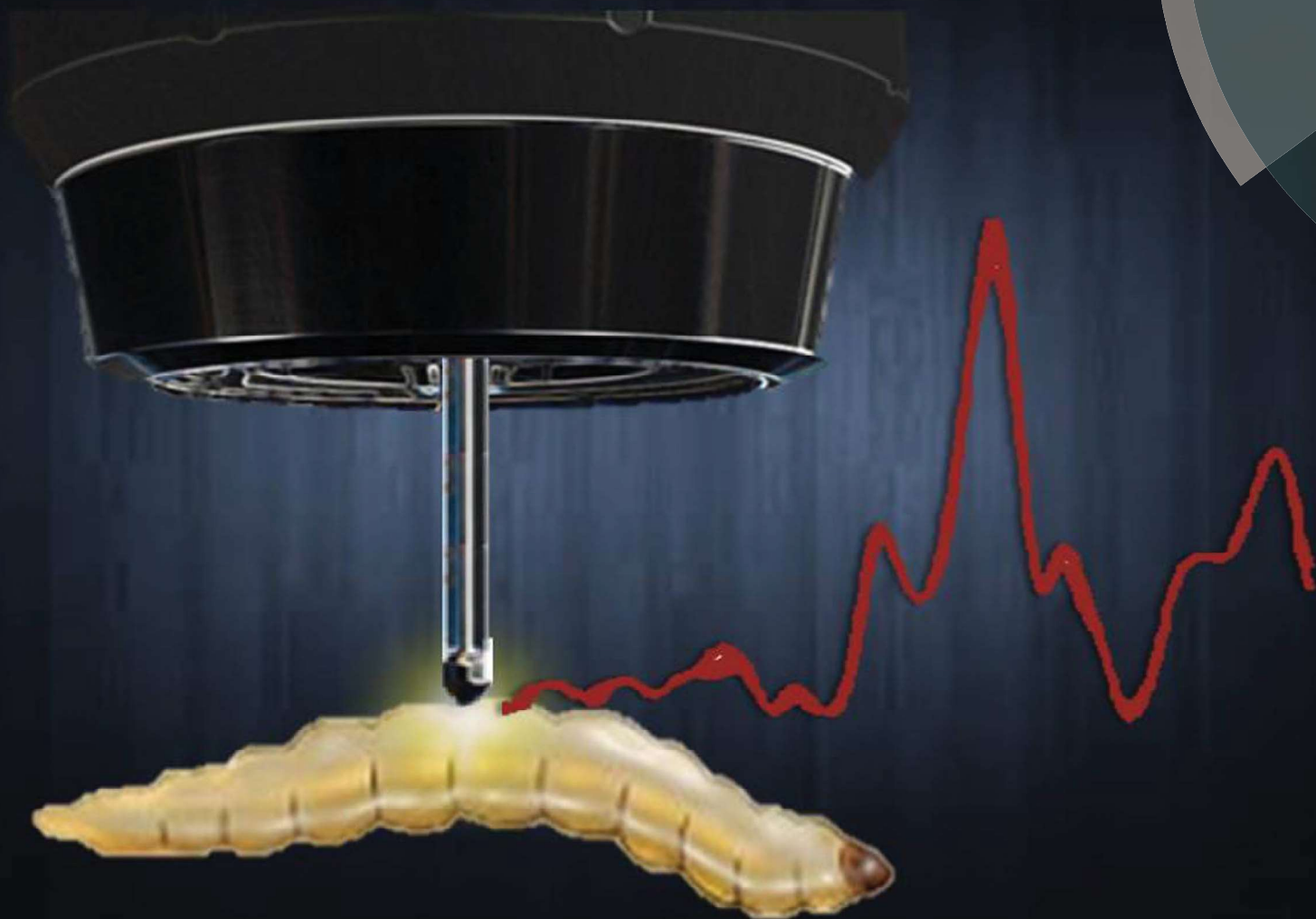
- Realizei a aquisição dos espectros;
- Realizei o processamento dos dados e construção dos modelos multivariados;
- Ajudei na escrita do manuscrito.

Leomir A. S. de Lima

Prof. Kássio M. G. Lima

Analytical Methods

www.rsc.org/methods



ISSN 1759-9660



PAPER

Kássio M. G. Lima *et al.*

FTIR microspectroscopy coupled with variable selection methods for the identification of flunitrazepam in necrophagous flies

175 YEARS

PAPER

Cite this: *Anal. Methods*, 2016, 8, 968

FTIR microspectroscopy coupled with variable selection methods for the identification of flunitrazepam in necrophagous flies

Tainá C. Baia,^a Renata A. Gama,^a Leomir Aires Silva de Lima^b and Kássio M. G. Lima^{*b}

The detection and identification of a drug in a corpse through the analysis of fly larvae feeding on the body by spectroscopic techniques promises to be of great value, because of their sensitivity, promptness, low cost and simplicity. Therefore, the purpose of this study was to develop a method based on Fourier-transform infrared (FTIR) microscopy to identify and discriminate flunitrazepam in necrophagous flies (*Chrysomya megacephala*, *Chrysomya albiceps* and *Cochliomyia macellaria*) as a non-invasive and non-destructive technique. Thirty-two Wistar mice were divided into two groups of sixteen and supplemented in two categories: group 1 – ethanol; and group 2 – standard flunitrazepam at a dose of 2 mg kg⁻¹. Spectra from the larvae samples were analyzed by principal component analysis-linear discriminant analysis (PCA-LDA), and variable selection techniques such as successive projection algorithm (SPA-LDA) and genetic algorithm (GA-LDA) to determine if control *versus* flunitrazepam could be segregated. In addition, the multivariate classification accuracy results were tested based on sensitivity, specificity, positive (or precision) and negative predictive values, Youden index, and positive and negative likelihood ratios. For control vs. flunitrazepam category, the sensitivity and specificity levels, using 46 wavenumbers by SPA-LDA, gave relatively good accuracy (up to 82.3% control vs. flunitrazepam). The resulting GA-LDA model also successfully classified both classes with respect to the main biochemical alterations induced by flunitrazepam using only 40 wavenumbers (up to 88.2% control vs. flunitrazepam). Compared to classical methods, this new approach could represent an alternative and an innovative tool for faster and cheaper evaluation in entomotoxicology.

Received 2nd September 2015

Accepted 3rd December 2015

DOI: 10.1039/c5ay02342d

www.rsc.org/methods

Introduction

Entomotoxicology ranks among the newest branches of forensic entomology that deals with the identification and quantitation of drugs and other toxins in carrion-feeding arthropods in decomposing tissues, and the study of drug-induced changes in arthropod growth with respect to the estimate of the post-mortem interval by entomological methods.^{1,2} The use of necrophagous species as a matrix for qualitative compound (drugs,^{3,4} metals^{5,6} and pesticides^{7,8}) detection in insect tissues has been generally accepted by forensic toxicologists. Although there has been recent progress in the detection of toxic substances in intact insects, there are some limitations such as insufficient knowledge of insect development and activity, proper use and validation of analytical procedures and lack of a general consensus concerning experimental set-up and sampling.

Several analytical drug detection/quantification procedures have been used for the analysis of insect tissues. These include radioimmunoassay,^{9,10} gas chromatography/mass spectrometry,^{11,12} and high-performance liquid chromatography-mass spectrometry,^{13,14} coupled with classic extraction techniques such as protein precipitation, liquid-liquid extraction or solid phase extraction. However, these techniques carried out in an entomotoxicological context are expensive, invasive, destructive, involve numerical preparation steps and most of the time they require pools of specimens to detect any present drug. As a consequence, there has been increased interest in the use of alternative, new methods for detection and identification of a drug being present in a corpse.

The detection and identification of a drug present in a corpse through the analysis of fly larvae feeding on the body by spectroscopic techniques promises to be of great value, because of their sensitivity, promptness, low cost and simplicity. Recently, we used near-infrared spectroscopy (NIRS) as a novel non-destructive method for the identification of flunitrazepam in *Chrysomya megacephala* (Fabricius) (Diptera: Calliphoridae) larvae, puparia and adults,¹⁵ with the resulting study successfully detecting biochemical alterations for the insect. NIRS is characterized by broad overlapping spectral peaks produced by

^aDepartment of Microbiology and Parasitology, Federal University of Rio Grande do Norte, Natal 59072-970, RN, Brazil^bInstitute of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio Grande do Norte, Natal 59072-970, RN, Brazil. E-mail: kassiolima@gmail.com; Tel: +55 84 3342 2323

the overtones and combinations of infrared vibrational modes. While these results are encouraging, other databases of vibrational spectra for entomotoxicological results must be established. Fourier-transform infrared (FTIR) spectroscopy is one technique with potential applications in the field of entomotoxicology. FTIR spectroscopy has the ability to rapidly generate a “biochemical cell-fingerprint” of the material under analysis.¹⁶ IR spectroscopy is also characterized by a minimum of sample handling, requiring no extractions and is non-destructive.¹⁷ Unlike conventional techniques used in the analysis of insect tissues, IR spectroscopy yields a precise representation of all the chemical bonds present in a sample and offers the opportunity to very quickly observe all metabolic modifications induced by a specific compound.¹⁸ In this context, it could be interesting to develop a strategy based on IR spectroscopy for detection and identification of a drug in a corpse.

Also, the use of appropriate chemometric tools for multivariate calibration and classification is largely responsible for advancing spectroscopic techniques, for instance, IR and NIR. Computational approaches [e.g., principal component analysis (PCA),¹⁹ linear discriminant analysis (LDA),²⁰ genetic algorithm (GA)²¹ and successive projections algorithm (SPA)²²] permit the processing of large amounts of spectroscopic data variables that subsequently require data reduction approaches in order to identify sources of variance across spectra and for inter-class variation to be identified.

Not only is the choice and development of computational approaches important to ensure reliable drug detection and quantification using spectroscopic techniques, but also multivariate classification quality features such as sensitivity, specificity, positive (or precision) and negative predictive values, Youden index, and positive and negative likelihood ratios should be calculated to ensure the validity of the results in accordance with international guidelines.²³ When reviewing entomotoxicological publications, the methods tend to lack proper validation.

Herein, we have attempted to evaluate the potential of a novel non-destructive method based on attenuated total reflection (ATR)-FTIR microspectroscopy for identification of flunitrazepam in 32 larvae. Flunitrazepam is the *N*-methyl-2'-fluoro analogue of nitrazepam and is available in a number of western European countries for use as a hypnotic (Rohypnol and Noriel®) and anesthetic (Narcozep®) agent. The detection of flunitrazepam, the most frequently abused pharmaceutical drug in the world, in necrophagous flies (*Chrysomya megacephala*, *Chrysomya albiceps* and *Cochliomyia macellaria*) as a non-invasive and non-destructive technique does not appear to be well documented. In our study, sample preparation, spectroscopic measurement, data preprocessing, feature selection and analytical validation were addressed. To our knowledge, there is no reported use of FTIR microscopy for the detection and identification of a drug being present in a corpse.

Materials and methods

Thirty-two Wistar mice (*Rattus norvegicus*) with an average weight of 300 g were divided into two groups of sixteen and

supplemented in the following way: group 1 – ethanol; and group 2 – standard flunitrazepam at a dose of 2 mg kg⁻¹. One hour after the supplementation, the mice were sacrificed, individually housed in fly traps and distributed at eight points along a track of a nearby forest. On the third and fourth days, 10 larvae were collected from each mouse and analyzed with ATR microspectroscopy. The average weight of one larva was estimated at 80 mg. All experiments were performed in compliance with the relevant laws and institutional guidelines, where the ethics committee of the Federal University of Rio Grande do Norte (UFRN/Brazil) have approved the experiments [Research and Ethics Committee (REC) approval no. 044/2013].

IR spectra [$n = 264$, 32 larvae (*Chrysomya megacephala*, *Chrysomya albiceps* and *Cochliomyia macellaria*) and eight random points] were collected from larvae (placing larvae individually on their backs on the plate) using a Bruker Lumus FTIR spectrometer with motorized ATR crystal (Bruker Optics Ltd, Coventry, UK). Prior to analyzing each specimen, the diamond crystal within the spectrometer was washed and a background spectrum was obtained to account for atmospheric composition.

The data import, pre-treatment and construction of chemometric classification models (PCA-LDA, SPA-LDA and GA-LDA) were implemented in MATLAB R2014a software (<http://www.mathworks.com>). Raw spectra were pre-processed by cutting between 1800 and 900 cm⁻¹ (235 wavenumbers at 3.84 cm⁻¹ spectral resolution) and baseline-corrected. For PCA-LDA, SPA-LDA and GA-LDA models, the samples were divided into training (70%), validation (15%) and prediction sets (15%) by applying the classic Kennard–Stone (KS) uniform sampling algorithm to the IR spectra. The KS algorithm was applied separately to each class to extract a representative set of objects from a given dataset by maximizing the minimal Euclidean distance between already selected objects and the remaining objects. The training samples were used in the modelling procedure (including variable selection for LDA), whereas the prediction set was only used in the final evaluation of the classification. The optimum number of variables for SPA-LDA and GA-LDA was determined with an average risk G of LDA misclassification. Such a cost function is calculated in the validation set as:

$$G = \frac{1}{N_V} \sum_{n=1}^{N_V} g_n, \quad (1)$$

where g_n is defined as

$$g_n = \frac{r^2(x_n, m_{I(n)})}{\min_{I(m) \neq I(n)} r^2(x_n, m_{I(m)})} \quad (2)$$

where $I(n)$ is the index of the true class for the n th validation object x_n .

In this definition, the numerator is the squared Mahalanobis distance between object x_n (of class index I_n) and the sample mean $m_{I(n)}$ of its true class. The denominator in eqn (2) corresponds to the squared Mahalanobis distance between object x_n and the center of the closest wrong class.

The GA routine was carried out during 40 generations with 80 chromosomes each. Crossover and mutation probabilities were set to 60% and 10%, respectively. Moreover, the algorithm was repeated three times, starting from different random initial populations. The best solution (in terms of the fitness value) resulting from the three realizations of the GA was employed. For this study, LDA scores, loadings, and discriminant function (DF) values were obtained for the specimen.

Sensitivity (the confidence that a positive result for a sample of the label class is obtained), specificity (the confidence that a negative result for a sample of non-label class is obtained), positive predictive value (PPV; measures the proportion of correctly assigned positive examples and its value varies between 0 and 1), negative predictive value (NPV; measures the proportion of correctly assigned negative examples and its value varies between 0 and 1), Youden's index (YOU; evaluates the classifier's ability to avoid failure), and the likelihood ratios (LR+) (represents the ratio between the probability of predicting an example as positive when it truly is positive, and the probability of predicting an example as positive when it actually is not positive) and (LR−) (represents the ratio between the probability of predicting an example as negative when it is actually positive, and the probability of predicting an example as negative when it truly is negative) were calculated as important quality standards in test evaluation. The quality metrics used in this study for evaluating the classification results can be calculated using the following equations:

$$\text{Sensitivity}(\%) = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100 \quad (3)$$

$$\text{Specificity}(\%) = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100 \quad (4)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (6)$$

$$\text{YOU} = \text{SENS} - (1 - \text{SPEC}) \quad (7)$$

$$\text{LR}(+) = \frac{\text{SENS}}{1 - \text{SPEC}} \quad (8)$$

$$\text{LR}(-) = \frac{1 - \text{SENS}}{\text{SPEC}} \quad (9)$$

where FN is defined as false negative, FP as false positive, TP as true positive and TN as true negative. SENS is defined as sensitivity and SPEC as specificity.

Results and discussion

In total, $n = 264$ spectra were acquired. The average IR spectrum for each original class (control, black line; flunitrazepam, red line) in the range 900–1800 cm^{-1} after baseline correction is shown in Fig. 1. As can be seen, discriminating between the two categories of specimens on the basis of IR measurements is not straightforward, owing to the complexity of the spectra.

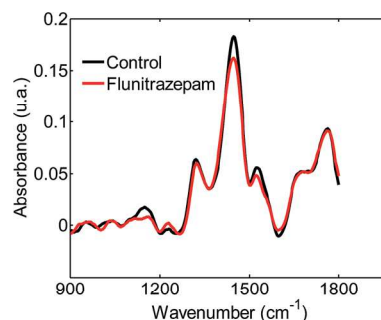


Fig. 1 Average spectrum for each original class (control, black line; flunitrazepam, red line).

Although all spectra had a similar shape in the regions 1150–1190 cm^{-1} , 1470–1490 cm^{-1} and 1505–1520 cm^{-1} , the spectra were shifted downwards.

Distinguishing these categories only by spectral observation is difficult; so, to identify markers, it is necessary to apply computational analysis (PCA-LDA, and variable selection techniques such as SPA-LDA and GA-LDA). The optimum numbers of PCs for PCA and variables for SPA-LDA and GA-LDA were determined by power *versus* cost calculation using the minimum cost function G . These were adopted to systematically classify normal *vs.* flunitrazepam based on IR spectra. Further, comparisons were made between rates, interpretability and training times.

Fig. 2 is a 2-D PCA-LDA Fisher scores plot of the derived spectral points from each category. We can see that the Fisher scores do not show good segregation. The PCA-LDA model was built using five PCs (93% variance in the data).

Then, SPA-LDA and GA-LDA were applied to the dataset to obtain the optimum number of variables by the minimum cost function G . Fisher scores for both models (SPA and GA) were obtained, and this improved the segregation between categories when compared with PCA-LDA.

SPA-LDA resulted in the selection of 46 variables (Table 1). Using solely 46 selected wavenumbers, the Fisher scores were calculated for both categories of the dataset, as shown in Fig. 3. As can be seen in Fig. 3, there is a greater effect of homogeneity among classes, with a little misclassification being obtained.

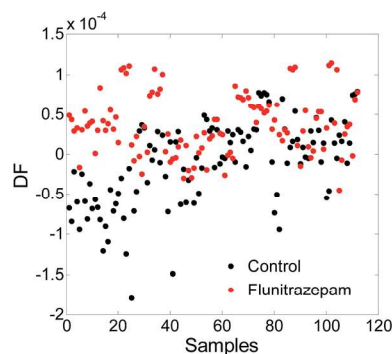


Fig. 2 Discriminant function *versus* samples calculated by using the PCA-LDA model from two categories (control and flunitrazepam).

Table 1 Variables for SPA-LDA and GA-LDA determined from the minimum cost function G used to achieve classification of control and flunitrazepam for a given validation dataset

Chemometric analysis	Wavenumber (cm^{-1}) selected							
SPA-LDA	1300	1304	1308	1315	1323	1327	1331	1339
	1345	1350	1355	1361	1367	1373	1378	1382
	1389	1396	1402	1409	1415	1420	1425	1430
	1434	1441	1445	1449	1454	1458	1462	1467
	1473	1478	1486	1492	1499	1505	1510	1517
	1522	1527	1534	1540	1545	1550		
GA-LDA	1304	1318	1319	1326	1327	1329	1331	1334
	1335	1342	1345	1352	1356	1369	1371	1382
	1386	1390	1391	1395	1403	1405	1421	1439
	1441	1445	1447	1448	1450	1451	1471	1472
	1477	1484	1486	1488	1500	1501	1505	1527

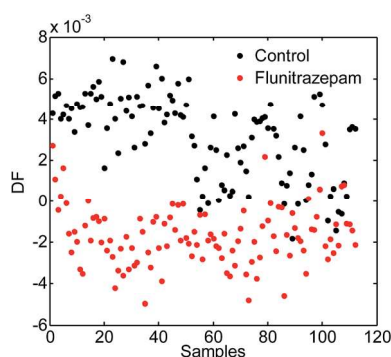


Fig. 3 Discriminant function versus samples calculated by the SPA-LDA model from two categories (control and flunitrazepam) using 46 wavenumbers selected.

Examination of the selected wavenumbers following SPA-LDA showed that the main biochemical alterations discriminating control vs. flunitrazepam were lipid, protein, nucleic acid, carbohydrate and, to a lesser extent, DNA vibrations. Several selected wavenumbers appear to be of particular interest, namely, the variables at 1315 cm^{-1} , 1389 cm^{-1} , 1505 cm^{-1} and 1550 cm^{-1} , associated with amide III of proteins, COO^- symmetric stretch in fatty acids, amide II of proteins and C–O stretching of predominantly α -sheet of amide II, respectively.

GA-LDA was applied to the dataset (control vs. flunitrazepam) and resulted in the selection of 40 variables (Table 1). Fig. 4 is a scores plot that shows that GA-LDA generates the best segregation between the two categories. Several selected wavenumbers (GA-LDA) appear to be of particular interest, namely, the variables at 1334 cm^{-1} and 1527 cm^{-1} , representing the amide III from proteins and CH bending and CH_2 wagging, respectively. These findings suggest that FTIR microscopy is a very promising technique for the non-destructive identification of flunitrazepam in *Chrysomya megacephala*, *Chrysomya albiceps* and *Cochliomyia macellaria* specimens. In addition, this finding is significant due to its potential for translation into entomotoxicology practice. Currently, no spectroscopic techniques are the gold standard in the detection procedures for the analysis of intact insects for toxicology decision-making.

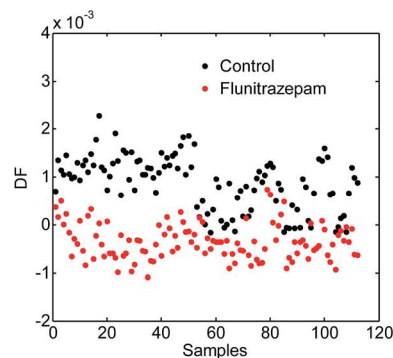


Fig. 4 Discriminant function versus samples calculated by the GA-LDA model from two categories (control and flunitrazepam) using 40 wavenumbers selected.

Classification rates were determined by using the best models. Table 2 presents the validation results for the optimized model (PCA-LDA, SPA-LDA and GA-LDA) of each category. According to results of sensitivity shown in Table 2, it is possible to see that the sensitivity rate from PCA-LDA, SPA-LDA and GA-LDA achieved scores of 64.7%, 70.5% and 64.7% for the control category, respectively, showing that the control category can be relatively well classified by these methods. For the flunitrazepam category, the sensitivity values from PCA-LDA, SPA-LDA and GA-LDA models were 64.7%, 82.3% and 88.2%, respectively. Furthermore, the specificity as shown in Table 2 for both categories suggests that following SPA-LDA and GA-LDA, an improved accuracy in comparison with PCA-LDA was obtained. In general, distinguishing between the control and flunitrazepam categories was more successful when using GA-LDA, demonstrating that ATR-FTIR microspectroscopy in conjunction with powerful chemometric approaches has the potential to detect and identify drugs present in a corpse.

Table 2 Values of quality performance features from three classification methods (PCA-LDA, SPA-LDA and GA-LDA) by FTIR microspectroscopy for each category

Stage performance features	PCA-LDA	SPA-LDA	GA-LDA
Control			
Sensitivity (%)	64.7	70.5	64.7
Specificity (%)	76.4	70.5	64.7
Positive predictive value (PPV)	73.3	70.5	64.7
Negative predictive value (NPV)	68.4	70.5	64.7
Youden index (YOU)	41.1	41.1	29.4
Positive likelihood ratio (LR+)	2.7	2.4	1.8
Negative likelihood ratio (LR–)	0.4	0.4	0.5
Flunitrazepam			
Sensitivity (%)	64.7	82.3	88.2
Specificity (%)	64.7	82.3	88.2
Positive predictive value (PPV)	64.7	82.3	88.2
Negative predictive value (NPV)	64.7	82.3	88.2
Youden index (YOU)	29.4	64.7	76.4
Positive likelihood ratio (LR+)	1.8	4.6	7.5
Negative likelihood ratio (LR–)	0.5	0.2	0.1

Conclusions

The results of this study show that ATR-FTIR microspectroscopy coupled with variable selection techniques (SPA or GA) of necrophagous flies may be an alternative approach for the detection and identification of flunitrazepam. We report a fast, clean and non-destructive methodology involving minimal sample preparation to categorize the specimens. In a case study for flunitrazepam samples, the resulting GA-LDA model successfully detected the biochemical alteration based on 40 wavenumbers and produced 82.3% and 88.2% sensitivity and specificity accuracies. For this dataset investigated, these selected wavenumbers (SPA-LDA and GA-LDA) appear to be of particular interest for the detection and identification of flunitrazepam. This is required to robustly validate the classification and biomarker extraction models for necrophagous flies and identification of flunitrazepam. Although the determination of abused drugs in insects is usually provided by gold standards (solid phase extraction chromatography methods), the proposed methodology can be applied to new drugs or necrophagous insects where the processing time and reagent costs required are a major advantage. This method was thoroughly validated in accordance with international guidelines, being considered suitable for use as an official methodology for entomological contexts. Further validation of these approaches exploiting other biospectroscopy techniques and using larger and architecturally more robust datasets is required.

Acknowledgements

The authors would like to thank PPGQ-UFRN for scientific support. K. M. G. Lima acknowledges the CNPq/Capes project (grants 070/2012 and 442087/2014-4), FAPERN (PPP 005/2012) for financial support. We are grateful to Fabio Godoy (Bruker Optics Ltd) for excellent technical assistance for this study with the Bruker Lumus FTIR spectrometer.

References

- 1 A. Tracqui, C. Keyser-Tracqui, P. Kintz and B. Ludes, *Int. J. Leg. Med.*, 2004, **118**, 194–196.
- 2 M. Gosselin, S. M. R. Wille, M. D. M. R. Fernandez, V. Di Fazio, N. Samyn, G. de Boeck and B. Bourel, *Forensic Sci. Int.*, 2011, **208**, 1–9.
- 3 C. P. Campobasso, M. Gherardi, M. Caligara, L. Sironi and F. Introna, *Int. J. Leg. Med.*, 2004, **118**, 210–214.
- 4 B. Levine, M. Golle and J. E. Smialek, *Am. J. Forensic Med. Pathol.*, 2000, **21**, 59–61.
- 5 E. M. Roeterdink, I. R. Dadour and R. J. Watling, *Int. J. Leg. Med.*, 2004, **118**, 63–70.
- 6 L. Lagoo, L. S. Schaeffer, D. W. Szymanski and R. W. Smith, *J. Forensic Sci.*, 2010, **55**, 624–632.
- 7 R. A. Rashid, K. Osman, M. I. Ismail, R. M. Zuha and R. A. Hassan, *Trop. Biomed.*, 2008, **25**, 184–190.
- 8 X. Liu, Y. Shi, H. Wang and R. Zhang, *Forensic Sci. Int.*, 2009, **192**, 14–18.
- 9 B. Bourel, G. Tournel, V. Hedouin, M. Deveaux and M. L. Goff, *Forensic Sci. Int.*, 2001, **120**, 127–131.
- 10 V. Hedouin, B. Bourel, L. Martin-Bouyer, A. Becart, G. Tournel, M. Deveaux and D. Gosset, *J. Forensic Sci.*, 1999, **44**, 351–353.
- 11 C. P. Campobasso, M. Gherardi, M. Caligara, L. Sironi and F. Introna, *Int. J. Leg. Med.*, 2004, **118**, 210–214.
- 12 D. T. Manhoff, I. Hood, F. Caputo, J. Perry, S. Rosen and H. G. Mirchandani, *J. Forensic Sci.*, 1991, **36**, 1732–1735.
- 13 D. W. Sadler, J. Richardson, S. Haigh, G. Bruce and D. J. Pounder, *Am. J. Forensic Med. Pathol.*, 1997, **18**, 397–403.
- 14 D. W. Sadler, L. Robertson, G. Brown, C. Fuke and D. J. Pounder, *J. Forensic Sci.*, 1997, **42**, 481–485.
- 15 J. S. Oliveira, T. C. Baia, R. A. Gama and K. M. G. Lima, *Microchem. J.*, 2014, **115**, 39–46.
- 16 M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner and F. L. Martin, *Nat. Protoc.*, 2014, **9**, 1771–1791.
- 17 A. L. Mitchell, K. B. Gajjar, G. Theophilou, F. L. Martin and P. L. Martin-Hirsch, *J. Biophotonics*, 2014, **7**, 153–165.
- 18 A. Derenne, M. Verdonck and E. Goormaghtigh, *Analyst*, 2012, **137**, 3255–3264.
- 19 M. Cavagna, R. Dell'Anna, F. Monti, F. Rossi and S. Torriani, *J. Agric. Food Chem.*, 2010, **58**, 39–45.
- 20 L. F. Chiu, P. Y. Huang, W. F. Chiang, T. Y. Wong, S. H. Lin, Y. C. Lee and D. Bin Shieh, *Anal. Bioanal. Chem.*, 2013, **405**, 1995–2007.
- 21 M. Ferrand, B. Huquet, S. Barbey, F. Barillet, F. Faucon, H. Larroque, O. Leray, J. M. Trommenschlager and M. Brochard, *Chemom. Intell. Lab. Syst.*, 2011, **106**, 183–189.
- 22 A. S. Marques, M. C. N. de Melo, T. A. Cidral and K. M. G. de Lima, *J. Microbiol. Methods*, 2013, **98**, 26–30.
- 23 M. Thompson, S. L. R. Ellison and R. Wood, *Pure Appl. Chem.*, 2002, **74**, 835–855.

Apêndice C

MCR-ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel-diesel blends.

Anne B.F. Câmara

Luciene S. de Carvalho

Camilo L.M. de Moraes

Fernanda M. de Oliveira

Heloise O.M. de Araújo

Leomir A. S. de Lima

Kássio M.G. de Lima

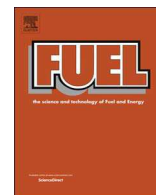
Fuel 210 (2017) 497–506

Contribuição:

- Realizei a aquisição dos espectros;
- Realizei o processamento dos dados e construção dos modelos MCR-ALS;
- Ajudei na escrita do manuscrito.

Leomir A. S. de Lima

Prof. Kássio M. G. Lima



Full Length Article

MCR-ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel-diesel blends



Anne B.F. Câmara, Luciene S. de Carvalho*, Camilo L.M. de Moraes, Leomir A.S. de Lima,
Heloise O.M. de Araújo, Fernanda M. de Oliveira, Kássio M.G. de Lima

Institute of Chemistry, Federal University of Rio Grande do Norte, Natal 59078-900, Brazil

ARTICLE INFO

Keywords:

Biodiesel/diesel blends
Fuel adulteration
FT-MIR
FT-NIR
MCR-ALS
PLS

ABSTRACT

Fourier transforming near infrared (FT-NIR) and mid infrared (FT-MIR) spectroscopies combined with multivariate curve resolution alternating least squares (MCR-ALS) and partial least squares (PLS) regression were utilized to identify and quantify residual vegetable oil adulteration in commercial diesel. In addition, physicochemical methods following ASTM standards were performed for comparison in identifying diesel adulteration. In total, 16 commercial diesel samples containing 8% biodiesel and 10 mg sulfur/kg (named S10B8) were prepared with the addition of residual vegetable oil (1–60%) and were analyzed by FT-NIR and FT-MIR spectroscopies. The physical–chemical properties analyzed were atmospheric distillation, flash point, kinematic viscosity and specific mass tests. Only the distillation curve was able to identify the adulteration. A paired t-test and an F-test were performed for the MCR-NIR, MCR-MIR, PLS-NIR and PLS-MIR models. Through the results, it can be stated that all the models used in this work are valid for a 95% confidence level. Furthermore, no statistical difference was observed between the estimated concentrations of adulterant and the reference values. In this work, MCR-ALS was applied as a spectral decomposition tool as a new approach to identify the adulterant profile on biodiesel/diesel blends, followed by its quantification. MCR-ALS was able to recover the related pure spectral profile of the fuels and adulterant, and to accurately predict the concentration of frying oil in biodiesel/diesel samples.

1. Introduction

In Brazil, the prevalent commercialized fuel is a blend of diesel with (8.0 ± 0.5) (%v/v) biodiesel [1,2]. These blends are usually subject to adulteration with the addition of cheaper raw materials such as kerosene [3,4], alcohol, solvents, vegetable oils [5,6] and waste cooking oil [7]. The illegal addition of cooked vegetable oil to diesel is one of the most usual adulterations due to its good miscibility and availability as a residual material [8]. The presence of adulterants in the fuel can cause financial losses, problems to the engine performance and environmental impact associated with corrosion, thereby increasing fuel consumption as well as emissions of particulate material and exhaust gases [9,10].

The quality of Brazilian commercial diesel is assessed by the use of different physicochemical parameters in accordance with resolution No. 30 of the National Agency of Petroleum Natural Gas and Biofuels (ANP) in the Fuel Quality Monitoring Program (PMQC), employing methodologies determined by ASTM International standards [11–13]. Flash point [14], distillation [15], kinematic viscosity [16], specific mass [17] and cetane number [18] are some of the physicochemical

parameters used to monitor the quality of diesel and ensure appropriate burning in the engines [11,19].

However, these analyses are time-consuming and require relatively high cost equipment. They also generate polluting emissions and residues; furthermore, these parameters are not sufficient to establish if an adulteration has occurred [2]. So, it is important to develop reliable analytical methods to evaluate the presence of contaminants in diesel fuel blends with the use of low cost instrumentation [19].

When associated with chemometric methods, vibrational spectroscopy (IR, NIR and Raman) has proven to be a powerful tool in analyzing fuel blends [2,20,21]. These spectroscopy analysis methods are non-destructive, present good accuracy and precision, and can be used in remote quality control [2,22], in addition to being faster than the usual ASTM methods.

Several chemometric methods have been applied to spectral data for diesel blends analysis. Multivariate calibration methods such as Multivariate Linear Regression (MLR) [23], Partial Least Squares (PLS) regression [5,9,24], Support Vector Machines (SVM) [25], and Artificial Neural Networks (ANN) [26] have been used to extract information

* Corresponding author.

E-mail address: luciene_car@hotmail.com (L.S. de Carvalho).

<http://dx.doi.org/10.1016/j.fuel.2017.08.072>

Received 25 May 2017; Received in revised form 22 July 2017; Accepted 18 August 2017
0016-2361/ © 2017 Elsevier Ltd. All rights reserved.

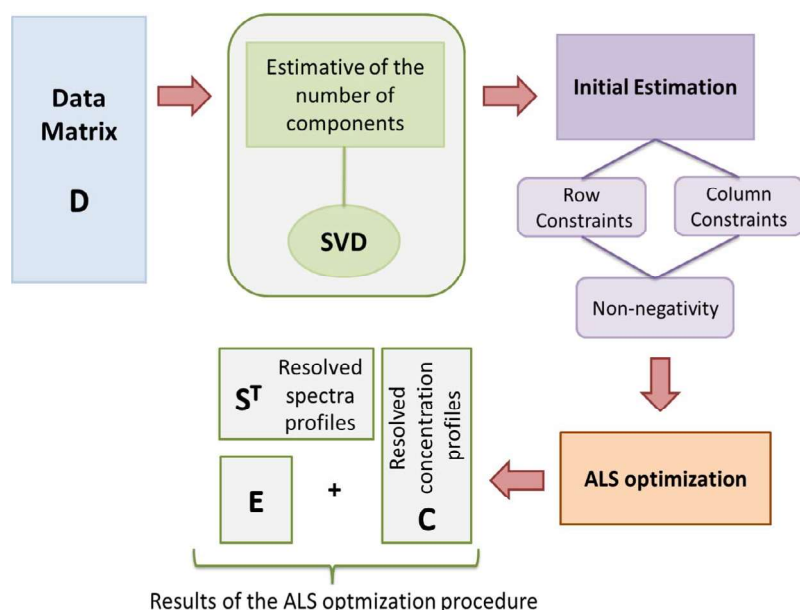


Fig. 1. Description of the MCR-ALS process.

Table 1
Physicochemical characterization data of frying oil, biodiesel and S10B8.

Property	Wasted frying oil	Biodiesel	S10B8
Kinematic viscosity at 40 °C (mm ² /s)	32.88	4.329	3.14
Specific mass at 20 °C (g/cm ³)	0.922	0.879	0.839
Water content (%m/m)	0.005	0.015	0.013

Table 2
Spectral data of NIR and FTIR of wasted frying oil and biodiesel.

Adulterant and biodiesel samples*			
FT-MIR spectra		FT-NIR spectra	
Wavenumber (cm ⁻¹)	Absorption feature	Wavelength (nm)	Absorption feature
1737	$\nu(\text{C}=\text{O})$	1737	C–H/C=O combination stretch
1160	$\nu(\text{C}-\text{O})$	1160	C–H deformation band and C–H stretch/C–C stretch combination band)
710	$\rho_{\text{as}}(\text{CH}_2)$	1800	CH ₂ symmetric stretch 1st overtone
2926	$\nu_{\text{as}}(\text{CH}_2)$	2400	Combination C–H stretch/C–C stretch form CH
2952	$\nu_{\text{as}}(\text{CH}_3)$		
2851	$\nu_{\text{s}}(\text{CH}_2)$		
2359	$\nu(\text{C}-\text{N})$		
1461	$\delta(\text{CH}_2)$		

ν : axial deformation; δ : angular deformation; angular deformation in-plane; s: symmetric; as: asymmetric.

* The frying oil and biodiesel samples showed the same main bands for FT-MIR and FT-NIR spectral data.

from NIR and FTIR spectra and for obtaining quality parameters of this fuel [27]. Most studies report the use of these spectroscopic methods in conjunction with chemometric models to quantify biodiesel content in diesel. A few of these papers present the quantification of adulterants in biodiesel/diesel blends with infrared spectroscopy [1,2,5,9,24,28–32], mainly using the PLS as a chemometric tool, but the identification of pure adulterant is not possible.

Some papers on biodiesel/diesel quality parameters are found in

literature using multivariate exploratory analysis and calibration methods, but only a few used multivariate resolution techniques such as multivariate curve resolution with alternating least squares (MCR-ALS) [27,33,34]. MCR-ALS decomposes a data set of multivariate mixed measurements in a bilinear model of pure component contributions, extracting a pure spectra of the sample constituents and related concentration profiles from the information contained in raw measured spectra [35,36]. MCR-ALS has been proven to be efficient in solving and providing relative determination of complex processes and mixtures [35] such as liquid chromatography with diode array detection for exploration, resolution and quantification of multicomponent systems [37]; EEM fluorescence data to determine cholesterol [38]; and NIR and UV–Vis for determining interesting compounds in biodiesel blends [27]. According to our research in specialized literature, MCR-ALS was not used with infrared spectroscopy in any work in this area. Recently, only one paper reported its application in the study of fuel adulteration, using Raman to extract the spectral data from the samples [39].

In this study Multivariate Curve Resolution (MCR) with alternating least squares (ALS) was used as a new approach, and optimization in conjunction with FT-NIR and FT-MIR spectroscopies for identifying and quantifying biodiesel/diesel blend (S10B8) adulteration by vegetable frying oil. In addition, MCR-ALS enabled the recovery of spectral profiles from the pure components of the blends, separating and identifying the contaminating substance, providing further investigation of the adulteration process. PLS regression was used as a comparative method to this new methodology for quantifying the adulterant, but it is not able to extract the spectral profile of the contaminant in the mixture.

2. Material and methods

2.1. Material

Samples of diesel S10 (diesel sample with 10 mg sulfur/kg) from the Potiguar Clara Camarão Refinery (RPCC) in Rio Grande do Norte – Brazil, were used in this work for analysis. A set of 16 different blends containing 8% (v/v) of soybean oil and animal fat biodiesel mixture (also from RPCC) in diesel were prepared to produce diesel S10B8. Samples of waste soybean frying oil were collected after its use in domestic food frying processes and an adulteration process of diesel/biodiesel mixtures was used. The samples and the contaminated mixtures were stored in appropriate polyethylene bottles, sealed and kept at room temperature until testing.

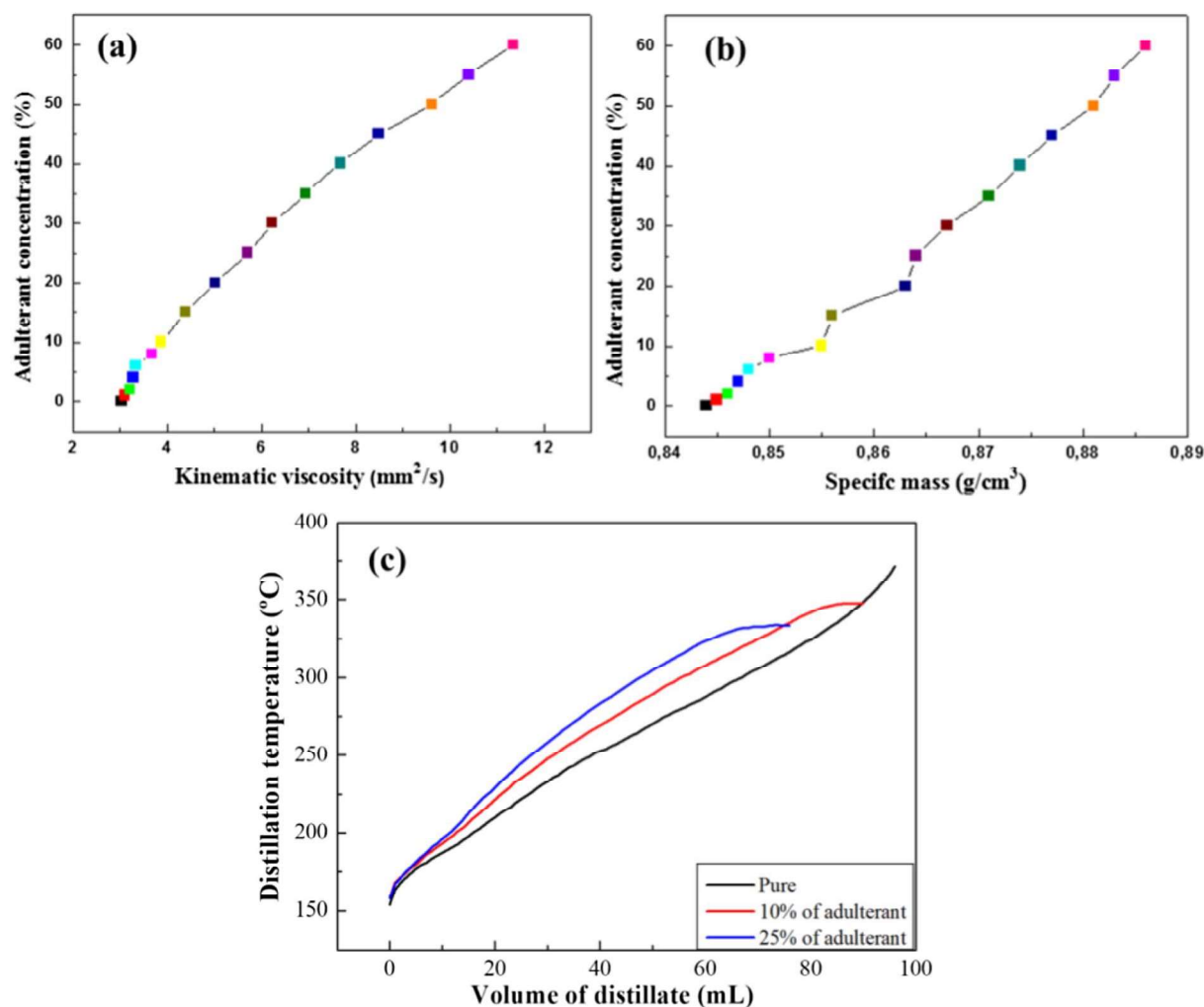


Fig. 2. Results for (a) specific mass; (b) kinematic viscosity tests and (c) Distillation curves for pure biodiesel-diesel and its blends with 10% and 25% (v/v) residual oil.

Table 3

Results of distillation and flash point for diesel S10B8 and its samples with 10% and 25% of residual oil.

Samples	Analysis					Flash point (°C)
	Atmospheric distillation					
	10% recovered (°C)	50% recovered (°C)	85% recovered (°C)	90% recovered (°C)	95% recovered (°C)	
Specification value [11]	180 (min)	245–295	note	note	note	38 (min)
Diesel S10B8	187.0	270.3	335.3	348.3	366.3	54
10% (v/v) residual oil	193.1	289.3	347.2	348.0	–	57
25% (v/v) residual oil	195.9	305.0	–	–	–	59

2.2. Experimental procedure

2.2.1. Samples preparation

Diesel S10B8 blends were prepared by adding biodiesel to diesel (92% diesel and 8% biodiesel). Before being used as contaminants, the soybean frying oil was filtered on cotton with a simple funnel in order to remove particulate matter. In order to obtain the mixtures of contaminated fuel, the corresponding volume of frying oil was added in concentrations ranging from 1% to 60% (v/v), with precision of 0.05%. For the chemometric models, 70% of the samples were used in the calibration set and 30% in the prediction set.

2.2.2. Physicochemical properties determination

In this work, fuel samples with and without adulterants were

submitted to physicochemical tests, following the methods described in ANP's resolution [11]. An Anton Paar Stabinger SVM-300 viscometer, a Petrotest's ADU 4+ automatic distillation unit and an automatic Pensky-Martens from Seta Mutiflash were used to determine specific mass and kinematic viscosity (ASTM-D7042) [16], atmospheric distillation (ASTM-D86) [15], flash point (ASTM-D93) [14] and water content (Karl Fischer method) (ASTM-E203) [40].

A viscodensimeter was used to perform viscosity and specific mass tests. Two milliliters (2 mL) of each fuel sample was transferred to the equipment cells using a clean and dry syringe. Kinematic viscosity measurement was performed at 40 °C in a cell containing a pair of rotating concentric cylinders, and specific gravity was performed at 20 °C in a U-shaped oscillator tube, where a variation in the oscillation frequency caused by a change in the tube mass was used in conjunction

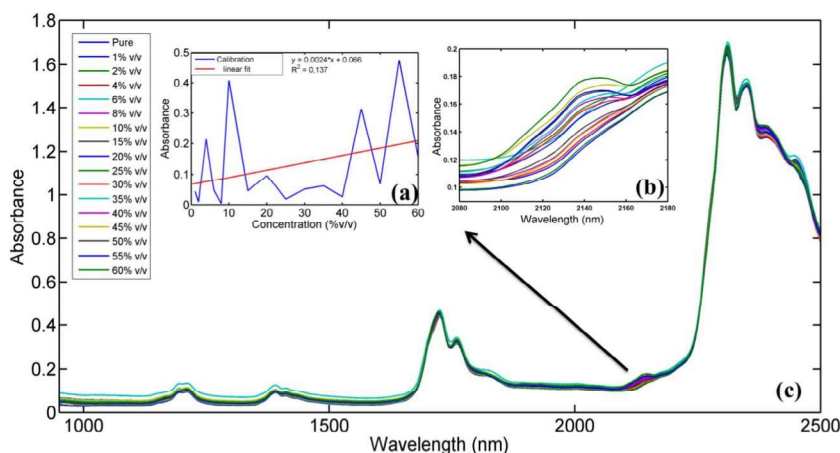


Fig. 3. FT-NIR spectra for the biodiesel-diesel blend and its adulterated samples.

with the calibration data for determining properties. The precision used for this method was 0.1%.

For the distillation process, 100 mL of sample was transferred to a specific distillation flask with a temperature sensor, and then heated to keep the distillation ratio between 4 and 5 mL/min. The sample vapor was condensed and collected in a beaker, and the volume recovered was measured with a digital sensor. The distillation curves (distillation temperatures according to the volume recovered) were obtained after correcting the temperature readings to atmospheric pressure of 101.3 kPa, and measuring the waste volume following ASTM-D86.

In order to achieve the flash point according to ASTM-D93, 70 mL of each sample was transferred to the test cell equipped with a temperature sensor and stirred. An ignition source was applied to the test cell at regular intervals of 1 °C, starting at a temperature of 8 °C below the set value and 20 °C above it. The system stirring was only interrupted by the detection of flash point. In this test, the ignition source causes sample steam combustion, and the result is expressed in °C.

The water content in fuel samples and adulterant was determined using Karl Fischer Titration according to ASTM-E203. This method allows for measuring the amount of water in a sample using volumetric titration. A Volumetric Karl Fischer Titrator HI903 by Hanna Instruments was used in conjunction with Sigma Aldrich Hydranal Composite 5 titrant and a solution composed of dry Methanol ($\leq 0.01\%$ water) and Chloroform (1:3 v/v) as a solvent medium. Water content values are expressed in $\mu\text{L/mL}$.

2.2.3. Spectroscopy measurements

Fourier transform near infrared spectra (FT-NIR) of the mixtures were obtained using a spectrometer from ARCOptix with quartz cuvette of 1.00 mm in transmission mode. The FT-NIR readings were performed using ARCSpectro software version 1.6 with wavelength ranging from 1000 to 2500 nm and resolution of 8 cm^{-1} . Fourier transform mid infrared spectra (FT-MIR) were obtained using the equipment VERTEX 70 from Bruker coupled with an attenuated total reflectance (ATR) and wavenumbers ranging from 400 to 4000 cm^{-1} (4 cm^{-1} resolution and 16 scans). Readings were performed with OPUS software. All spectra were recorded at room temperature.

2.3. Chemometric analysis

2.3.1. Software

All data treatment and chemometric analyses were performed in a MATLAB® R2012b environment (MathWorks, USA) using PLS Toolbox version 7.9.3 (Eigenvector Research, Inc., USA) and MCR-ALS toolbox version 2.0 [41]. The samples were divided into calibration ($n = 11$) and prediction ($n = 5$) sets by using Kennard-Stone uniform sampling selection algorithm [42].

2.3.2. Preprocessing

FT-NIR measurements were carried out on a portable instrument, and the spectra were cut between 1000 and 2500 nm to remove ambient noise at higher wavelengths. In addition, transmittance FT-MIR spectra were transformed into absorbance scale before chemometric analysis. Both FT-NIR and FT-MIR spectra were preprocessed by Savitzky-Golay smoothing (window of 7 points) in order to remove noise.

2.3.3. Multivariate curve resolution alternating least squares (MCR-ALS)

Fig. 1 depicts the MCR-ALS model used for the identification and quantification of adulterant in fuels.

MCR is a bilinear model that is the multi-wavelength extension from Lambert-Beer's law [43–45], and can be described by Eq. (1),

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (1)$$

where \mathbf{D} ($n \times m$) is a data matrix containing the NIR or FTIR spectra of n samples for the m wavelengths recorded; \mathbf{C} $n \times A$ and \mathbf{S}^T $A \times m$ are the matrices with the concentration and spectra profiles of A pure components in the samples, respectively. \mathbf{E} has the same size of \mathbf{D} and contains the unexplained variance from the bilinear model, related as the experimental error [37,43,44,46].

Some constraints must be applied in each iteration to reduce the number of possible solutions for \mathbf{C} and \mathbf{S}^T , and to give chemical meaning to the results. Non-negativity constraint was applied in this work. This constraint forces the concentration and/or spectral profile to be equal or larger than zero [47].

2.3.4. Partial least square (PLS) regression

PLS regression is one of the most known techniques of multivariate calibration. It decomposes the spectra data \mathbf{X} and the concentration vector \mathbf{y} , as follows [48,49]:***

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E} \quad (2)$$

$$\mathbf{y} = \mathbf{U}\mathbf{q}^T + \mathbf{F} = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{F} \quad (3)$$

where \mathbf{T} is the score matrix of \mathbf{X} ; \mathbf{P} is the loading matrix of \mathbf{X} ; \mathbf{E} is the residual matrix of \mathbf{X} ; \mathbf{t}_h is the score of \mathbf{X} for a factor h ; \mathbf{p}_h is the loading of \mathbf{X} for a factor h ; \mathbf{U} is the score matrix of \mathbf{y} ; \mathbf{q} is the loading of \mathbf{y} ; \mathbf{F} is the residual of \mathbf{y} ; \mathbf{u}_h is the score of \mathbf{y} for each factor h ; and \mathbf{q}_h is the loading of \mathbf{y} for each factor h . Each component for PLS regression is obtained by maximizing the covariance between \mathbf{y} and all possible linear functions of \mathbf{X} . As a result, a set of factors (or latent variables, LVs) is obtained [50].

2.3.5. Statistical evaluation

In order to evaluate the prediction ability of the multivariate calibration models, a set of figures of merit were calculated in the

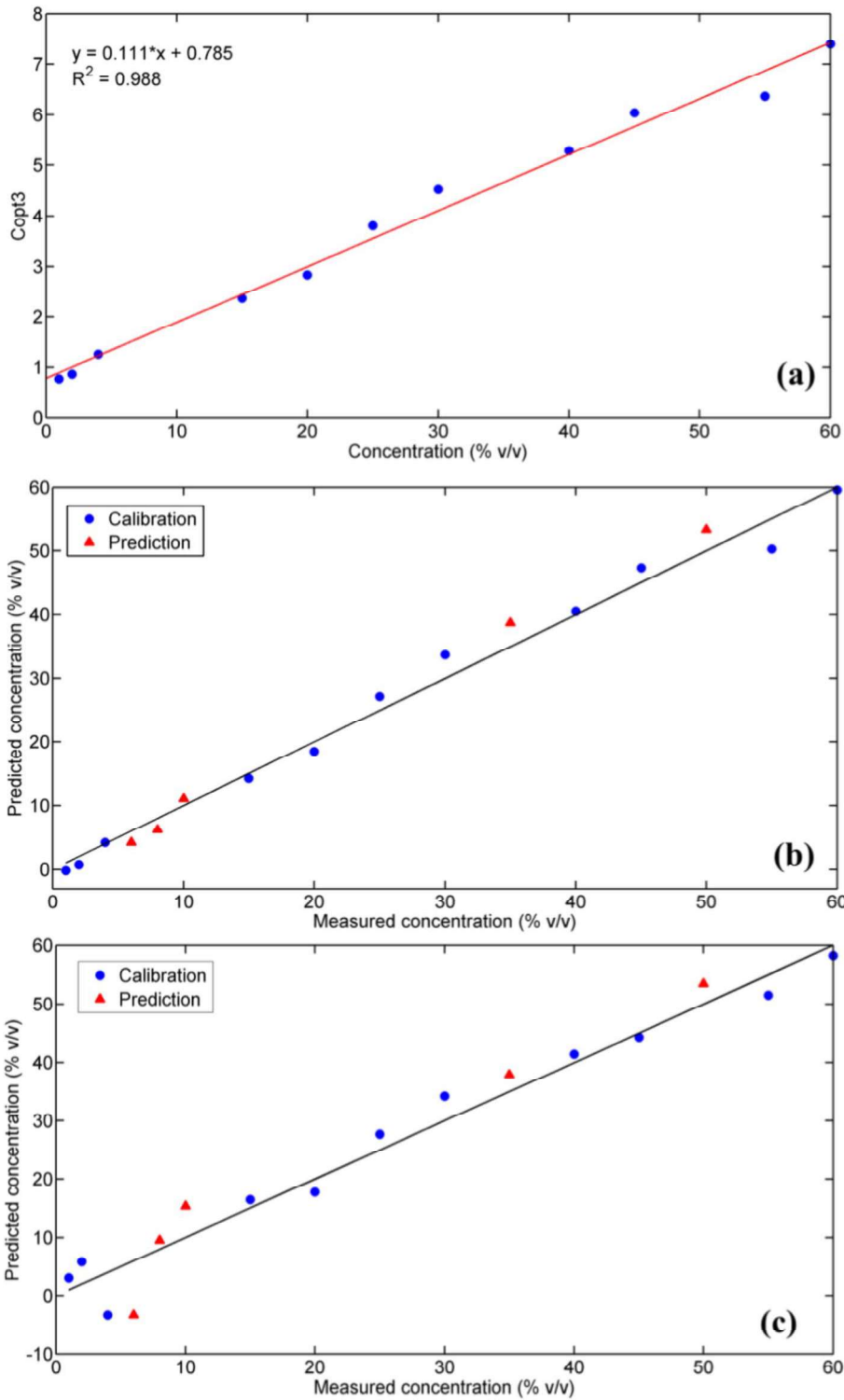


Fig. 4. Concentration profile for (a) Copt 3; and prediction curve for (b) MCR-ALS-NIR and (c) PLS-NIR.

predictions set. These figures were calculated according the following expressions [51]:

Root mean square error of prediction (RMSEP),

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

Standard error of prediction (SEP),

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \tag{5}$$

Bias,

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \tag{6}$$

Table 4
Figures of merit (FOM) for MCR-ALS and PLS models based on FT-NIR spectra.

FOM	MCR-ALS		PLS	
	Calibration	Prediction	Calibration	Prediction
Slope	0.997	1.118	0.971	1.128
Offset	0.004	−1.683	0.812	−2.000
R	0.994	0.998	0.986	0.974
R ²	0.988	0.996	0.973	0.948
Bias (% v/v)	0.07	−0.89	−0.02	−0.79
Relative error (%)	6.49	9.06	9.87	18.58
RMSE (% v/v)	2.18	2.54	3.32	5.20
SE (% v/v)	2.28	2.66	3.48	5.75

Relative percentage error in concentration prediction (RE%),

$$RE(\%) = 100 \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n y_i^2}} \quad (7)$$

where y_i e \hat{y}_i are the actual and predicted concentrations in sample i , respectively; and n is the total number of samples used in the prediction set.

3. Results and discussion

3.1. Characterization of frying oil, biodiesel and S10B8 samples

Preliminary characterization of the physicochemical data from the materials, residual soybean frying oil, biodiesel (RPCC) and diesel S10B8 used to obtain the test mixtures of this work, are presented in Table 1. In addition, the chemical characterization of these materials was performed by NIR and FTIR spectroscopies to obtain the spectral profile of these oils, whose main bands are described in Table 2.

As shown in Table 1, the viscosity of vegetable oil is much higher than that of biodiesel and diesel S10B8, and is initially a property capable of reporting diesel adulteration processes or diesel/biodiesel blends. The specific mass at 20 °C did not present a difference capable of identifying such adulterations due to the similar values obtained for the three materials. The samples of the three materials presented values of water content within the specification required by ANP Resolution No. 30 [11], which indicates a maximum value of 0.02% for diesel.

Table 2 shows FT-MIR data for biodiesel and residual vegetable oil, where the infrared absorption bands at 1737 cm^{−1} and 1160 cm^{−1} involving the C=O bonds that are present in both materials indicating that they have similarities in their chemical compositions due to the organic function of vegetable oil and biodiesel esters, since this is the

product of the transesterification reaction of the oils. Similar behavior is perceived for the NIR, where it can be seen that the two materials present bands in 2130 nm referring to the combination of C=O. Therefore, it is not possible to distinguish the presence of the two materials only using spectroscopic methods, and it is necessary to couple other tools such as chemometrics which contribute to the resolution of complex mixtures, thereby allowing such identification.

3.2. Physicochemical properties

Fig. 2a–c depict the kinematic viscosity profile at 40 °C, the specific mass profile at 20 °C and distillation curves for the analyzed samples, respectively. In observing these profiles, it is possible to observe a similar behavior for the properties, where a linear increase can be noticed as the concentration from the residual oil is increased. However, it is not possible to confirm the presence of adulteration below 10% (v/v) adulterant in the samples, since the values obtained for these parameters are still in accordance with the ANP specification [11].

Table 3 shows the values of flash point and distillation points for the commercialized diesel and two randomly chosen adulterated samples. These tests were not carried out for all the adulterated fuel blends because of the large volume of sample required and the long time for analysis. It is possible to observe an increase in flash point due to the incorporation of residual oil into diesel S10B8, but even with this increase, the blends remain in accordance with the specification for this parameter, where points above 38 °C are accepted even though a small increase is observed [11,52].

The distillation temperatures for 10%, 50%, 85%, 90% and 95% recovered volume, as specified by the ANP resolution, are presented in Table 3, where the adulterated samples were not according to the specification since they had an end point before the temperature of 95% distillate. As seen in Fig. 2c, it can be observed from the boiling point behavior of the samples during distillation that adulterated samples had higher boiling points than the commercialized diesel S10B8. In addition, the fuel blend with 25% (v/v) residual oil has an end point around 78% distillate, while the end point for the sample with 10% (v/v) contaminant occurred around 91% distillate. This fact occurs due to the compositional difference between the diesel and the vegetable oil [53].

Distillation would be the best physicochemical property technique for identifying adulteration, since changes in the curve profiles occur when chemical composition of the fuel changes. However, this technique has several disadvantages, such as the high analysis time and the damage that residual oil can inflict on the equipment.

3.3. Chemometric models

3.3.1. NIR data

Fig. 3 depicts the FT-NIR spectra for diesel S10B8 and all the

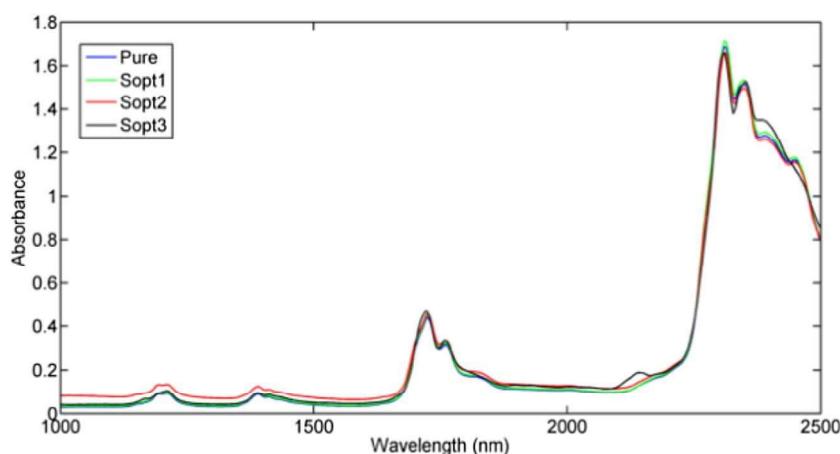


Fig. 5. Recovered Sopt profiles for FT-NIR data.

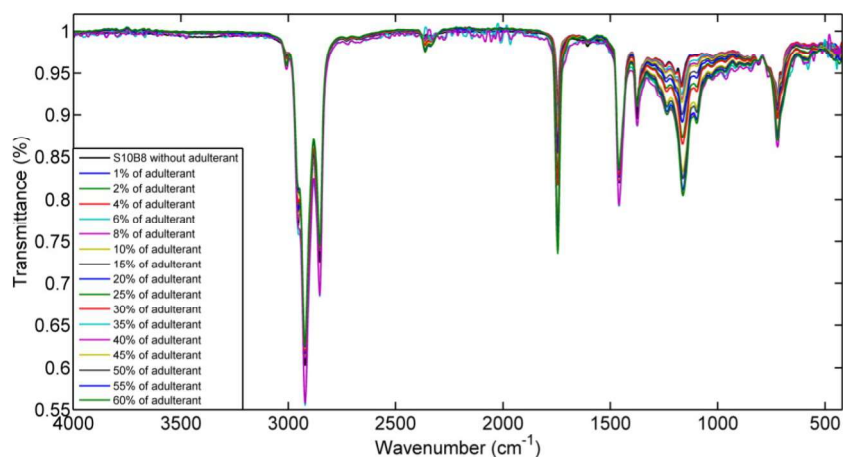


Fig. 6. FT-MIR spectra for the biodiesel-diesel blend and its adulterated samples.

contaminated samples, as well as amplification in the 2130 nm band and its linearity curve.

As shown in Fig. 3, spectral changes caused by the addition of adulterant to the fuel seem to be very small, and can mainly be seen in the 2130 nm band, corresponding to the combination of C–H/C=O stretch. An increase in intensity of that band can be seen as the concentration of the residual frying oil is increased, but this increase does not present a linear behavior, as can be seen in the linearity curve and on the amplification of this band. This could happen because NIR spectra have very superposed bands with relatively low signal-to-noise ratio [54].

MCR-ALS was applied to solve the spectra mixture in 3 components (explained variance of 99.998%, lack of fit of 0.394%). Among the 3 principal components solved by MCR-ALS, only the third (Copt3) had a linear relation with the increased adulterant concentration (Fig. 4a). The first and second PCs (Copt1 and Copt2) did not have any linear relation with concentration; in fact, their dependence on concentration seems to be random (as can be seen in the Supplementary material, Fig. 1). PLS was also applied to the data in order to build a calibration model. The measured *versus* predicted concentration by the third component of MCR-ALS and by PLS regression (LVs = 2, explained variance of 99.03%) are shown in Fig. 4b and c, respectively.

The PLS regression model has more residue, while MCR-ALS is better fit according to the prediction curves (Fig. 4). This is observed in the RMSEC, RMSEP and R^2 values, which were significantly smaller for MCR-ALS, showing that this model adjusted better to the dataset than PLS. Table 4 shows the figures of merit (FOM) calculated for both MCR-ALS and PLS models.

Despite both models having relatively close calibration performance with RMSEC and R^2 close to each other, the error in prediction set almost doubles in the PLS model, as demonstrated by the relative error, RMSEC and SEP values. This fact suggests that MCR-ALS was more selective for this data, in which only the 3rd component referent to the waste vegetable oil was enough to build the calibration curve (Copt3). In using both LVs in PLS, noise or other influences unrelated to adulterant concentrations possibly affected the model performance. Therefore, MCR-ALS was the best option to extract only the signal of interest from the mixture. Fig. 5 shows the MCR-ALS Sopt, which are the recovered spectral profiles of the blend components and have a linear dependence on their concentrations.

In Fig. 5, Sopt1 is the recovered profile from diesel and Sopt2 refers to the biodiesel spectra. The third Sopt is the most different since it corresponds to the waste cooking oil profile. In this Sopt, there are bands at ~2130 nm and ~2375 nm referent to combination C–H/C=O stretch and C–H deformation band and C–H stretch/C–C stretch combination band, respectively. In addition, there is band suppression at ~1800 nm, ~2400 nm and ~2450 nm referent to the 1st overtone of CH₂ symmetric stretch, combination C–H stretch/C–C stretch form

CH, and combination C–H stretch/C–C stretch form CH₂, respectively [55].

3.3.2. FT-MIR data

Fig. 6 depicts the FT-MIR spectra in transmittance mode. As the adulterant is added to the blends, it is possible to note the increase in intensity of some bands in the FT-MIR spectra. This difference can be noticed in the fingerprint region bands at 1737 and 1160 cm^{−1} that correspond to axial deformation from C=O of saturated aliphatic esters and the axial deformation of C–O, respectively [56].

It can be seen in Fig. 6 that the intensity of the feature at 1160 cm^{−1} has a linear tendency with the increase in waste frying oil concentration (see Supplementary Material, Fig. II); however, in the case of an analysis performed with real samples, it is not possible to state if the intensity of this band refers to the presence of frying oil or biodiesel, since it is present in the spectral profile of both substances, and the biodiesel concentration in commercial diesel is variable [11]. According to Guimarães and collaborators [28], quantitative analysis by infrared spectroscopy is only possible with the use of chemometric methods such as PLS, MLR and principal components regression (PCR). This is due to the fact that spectroscopic techniques do not resolve the components in a sample; chemical information about components is embedded in multiple bands in the spectra. Spectroscopic instruments alone provide very limited information for the unambiguous identification of unknown mixtures [57].

By applying MCR-ALS to the data, only two components were selected (explained variance of 99.886%, lack of fit of 3.376%). The first component has an inverse linear relation with the concentration of adulterant (Copt1, as can be seen in the Supplementary Material, Fig. III), while the second has a direct linear relation (see Fig. 7a). Therefore, the second component was chosen to build the calibration model using MCR-ALS.

In order to evaluate the performance of PLS regression, a PLS model was built using 2 LVs (99.19% explained variance). Measured versus predicted concentration using MCR-ALS and PLS regression are shown in Fig. 7b and c, respectively. The prediction performance of MCR-ALS and PLS were very similar to each other, in which the MCR-ALS had RMSEC of 2.98% (v/v) and PLS had RMSEC of 2.83% (v/v). The prediction performance was slightly better in the PLS model, with a RMSEP and SEP ~1% (v/v) smaller than MCR-ALS. Figures of merit (FOM) for both MCR-ALS and PLS models are shown in Table 5.

The better performance achieved using PLS probably occurred due to the greater sensitivity of FT-MIR, since the signal-to-noise ratio in this case is higher and the signals represent fundamental vibration modes, having less superposition than NIR data [54]. Fig. 8 depicts the MCR-ALS Sopt profile for FT-MIR.

Sopt1 and Sopt2 are the recovered profile of the biodiesel/diesel blend and the waste cooking oil, respectively. The MCR-ALS Sopt (see

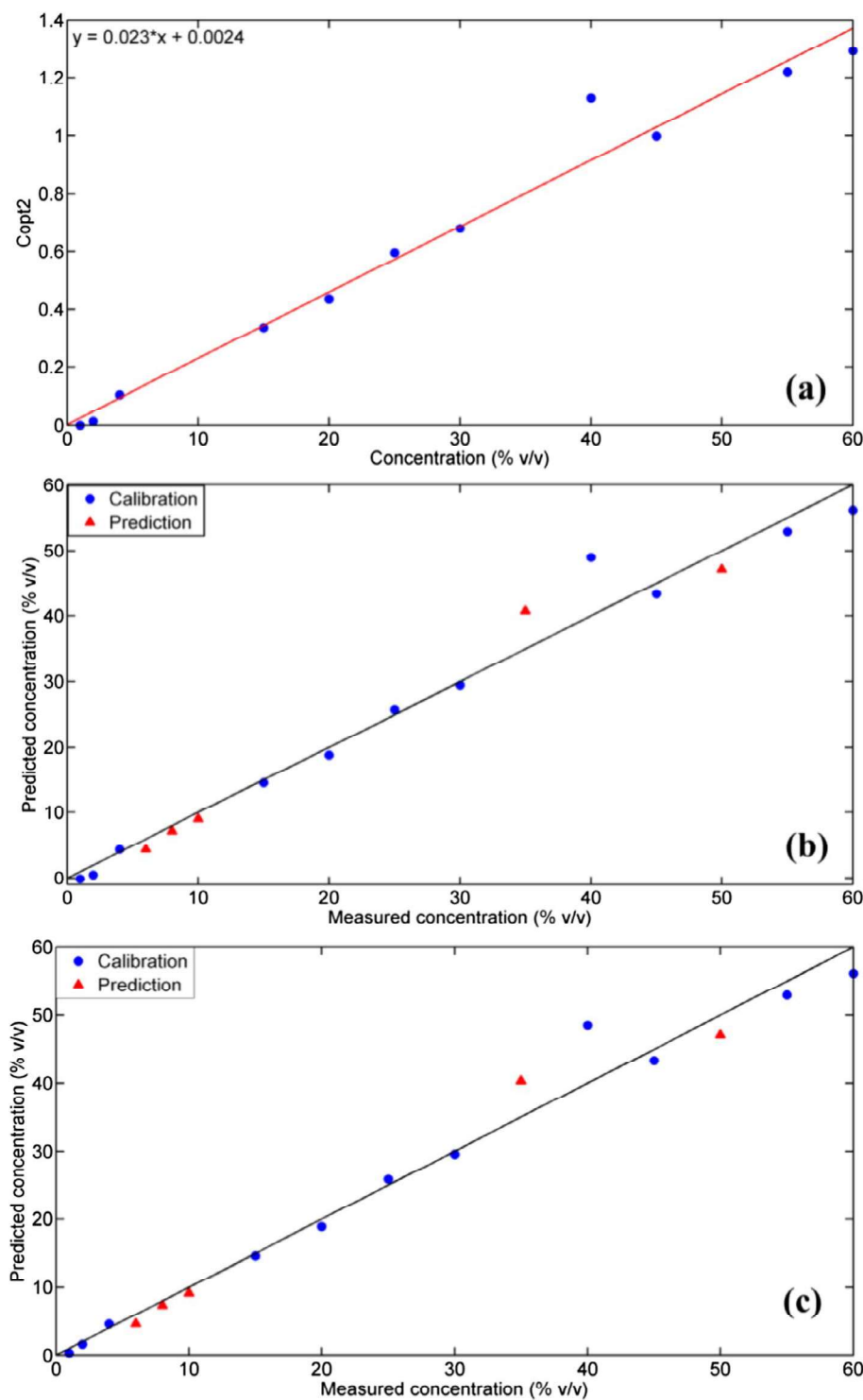


Fig. 7. Concentration profile for (a) Copt 2; and prediction curve for (b) MCR-ALS-MIR and (c) PLS-MIR.

Fig. 8) on the second component (Sopt2) shows higher intensity in the features at 1374 cm^{-1} , 1737 cm^{-1} and 710 cm^{-1} , referent to angular symmetric deformation of $\text{CH}_3\text{ C-H}$, C=O axial deformation of aliphatic and saturated esters and the angular asymmetric deformation in the CH_2 plane, respectively [56].

Some statistic tests were performed to determine the accuracy of the models. A paired t-test was performed, where t values of -0.748 , -0.462 , -0.475 and -0.476 were obtained for MCR-NIR, MCR-MIR, PLS-NIR and PLS-MIR, respectively. In comparing these values with the tabulated one (2.776), it is possible to say that the predicted concentration by both MCR-ALS and PLS models using FT-NIR and FT-MIR

data were statically equal to the reference ones within a confidence level of 95%. An F-test was also performed, where the values of 0.994, 0.923, 1.032 and 0.939 were obtained for MCR-NIR, MCR-MIR, PLS-NIR and PLS-MIR, respectively. As the value of Calculated $F < \text{Critical } F$ (Critical $F = 2.978$), it can be stated that all the models used in this work are valid for a 95% confidence level.

Finally, it is verified that some works in specialized literature use chemometric tools, especially PLS, in conjunction with NIR/FTIR spectroscopies for the quantification of adulterants in biodiesel blends with diesel [24,28–32]. However, the model applied in this study (MCR-ALS), has been shown to be efficient because in addition to

Table 5
Figures of merit (FOM) for MCR-ALS and PLS models based on FT-MIR spectra.

FOM	MCR-ALS		PLS	
	Calibration	Prediction	Calibration	Prediction
Slope	0.993	1.027	0.981	1.019
Offset	0.000	−0.684	0.461	−0.515
R	0.989	0.987	0.989	0.988
R ²	0.978	0.974	0.978	0.976
Bias (% v/v)	0.19	0.10	0.05	0.09
Relative error (%)	9.36	10.75	8.87	10.10
RMSE (% v/v)	3.14	3.01	2.98	2.83
SE (% v/v)	3.29	3.37	3.13	3.16

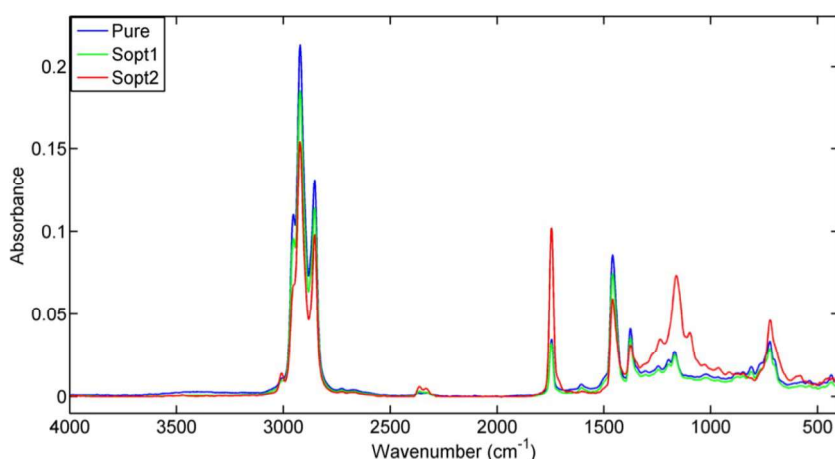


Fig. 8. Recovered Sopt profiles for FT-MIR.

identifying and quantifying adulterants, it is able to recover spectral profiles of these as pure components in the mixtures, allowing for identifying the contaminating substance and to provide greater investigation of adulteration in real samples. Using a database constructed with the spectra of possible commercial diesel adulterants, it will be possible to identify them in real adulterated samples through the Sopt profile extracted by the MCR-ALS.

4. Conclusion

Spectroscopic analysis combined with MCR-ALS and PLS chemometric methods were efficient in identifying and quantifying diesel adulteration and presented low errors (RMSEP < 5.56%) with correlation coefficients near 1. This indicates feasibility for using both FT-NIR and FT-MIR spectroscopies as complementary tools to measure adulteration in diesel blends, presenting low cost and good accuracy.

Physicochemical properties of diesel S10B8 and its contaminated blends were determined such as specific mass, kinetic viscosity and flash point, but they were not able to identify variation when the samples were adulterated with vegetable oil, according to ANP specification. On the other hand, it was possible to identify the adulteration with the atmospheric distillation curves even in small concentrations, but this technique has high analysis time, high cost and a great amount of sample is required.

Moreover, the innovation for this work was the use of MCR-ALS coupled with FTIR and FT-NIR spectroscopies, with the advantage of recovering the related pure spectral analyte and interference profiles, highlighting other chemometric tools, and can be used to identify several contaminants in biodiesel/diesel blends, thereby solving interference problems caused by only using spectroscopic techniques, while at the same time guaranteeing reliability in identifying and quantifying the adulteration.

Acknowledgements

The authors would like to acknowledge the Post-Graduate Chemistry Program PPGQ/UFRN, Oil and Derivatives Characterization Laboratory (LABPETROL), CAPES – Brazil and PRHPB-222 – Brazil for financial support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fuel.2017.08.072>.

References

- [1] Paiva EM, Rohwedder JJR, Pasquini C, Pimentel MF, Pereira CF. Quantification of biodiesel and adulteration with vegetable oils in diesel-biodiesel blends using portable near-infrared spectrometer. *Fuel* 2015;160:57–63.
- [2] Oliveira FCC, Brandão CRR, Ramalho HF, Costa LAF, Suarez PAZ, Rubim JC. Adulteration of diesel-biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT-Raman spectroscopy. *Anal Chim Acta* 2007;587:194–9.
- [3] Yadav ShR, Murthy KV, Mishra D, Baral B. Estimation of petrol and diesel adulteration with kerosene and assessment of usefulness of selected automobile fuel quality test parameters. *Int J Environ Sci Technol* 2005;4:253–5.
- [4] Cunha DA, Montes LF, Castro EVR, Barbosa LL. NMR in the time domain: a new methodology to detect adulteration of diesel oil with kerosene. *Fuel* 2016;166:79–85.
- [5] Vasconcelos FVC, Souza Jr PFB, Pimentel MF, Pontes MJC, Pereira CF. Using near-infrared overtone regions to determine biodiesel content and adulteration of diesel-biodiesel blends with vegetable oils. *Anal Chim Acta* 2012;716:101–7.
- [6] Basri KN, Hussain MN, Bakar J, Sharif Z, Khir MFA, Zoofakhar AS. Classification and quantification of palm oil adulteration via portable NIR spectroscopy. *Spectrochim Acta, Part A* 2017;173:335–42.
- [7] Corgozinho CNC, Pasa VMD, Barbeira PJS. Determination of residual oil in diesel oil by spectrofluorimetric and chemometric analysis. *Talanta* 2008;76:479–84.
- [8] Meira M, Quintella CM, Ferrer TM, Silva HRG, Guimarães AK, Santos MA. Identification of adulteration of biofuel by addition of residual oil instead of bio-diesel to the diesel by total spectrofluorimetry and principal component analysis.

- Quim Nova 2011;34:621–4.
- [9] Gaydou V, Kister J, Dupuy N. Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel-biodiesel blends by vegetal oil. *Chemometr Intell Lab Syst* 2011;106:190–7.
 - [10] Soares IP, Rezende TF, Pereira RCC, Santos CG, Fortes ICP. Determination of biodiesel adulteration with raw vegetable oil from ATR-FTIR data using chemometric tools. *J Braz Chem Soc* 2011;22:1229–35.
 - [11] ANP – Agência Nacional do Petróleo, Gás Natural e Biocombustíveis; Resolução No. 30 de 23.12.2013 – DOU 13.01.2017.
 - [12] Aleme HG, Corgozinho CNC, Barbeira PJS. Diesel oil discrimination by origin and type using physicochemical properties and multivariate analysis. *Fuel* 2010;89:3151–6.
 - [13] Aleme HG, Barbeira PJS. Determination of flash point and cetane index in diesel using distillation curves and multivariate calibration. *Fuel* 2012;102:129–34.
 - [14] ASTM D 93-13. Standard Test Methods for Flash Point by Pensky-Martens Closed Cup Tester. West Conshohocken (PA): ASTM International; 2013.
 - [15] ASTM D 86-12. Standard Test Method for Distillation of Petroleum Products at Atmospheric Pressure. West Conshohocken (PA): ASTM International; 2013.
 - [16] ASTM D 7042-14. Standard Test Method for Dynamic Viscosity and Density of Liquids by Stabinger Viscometer (and the Calculation of Kinematic Viscosity). West Conshohocken (PA): ASTM International; 2014.
 - [17] ASTM D4052-11. Standard Test Method for Density, Relative Density and API Gravity of Liquids by Digital Density Meter. West Conshohocken (PA): ASTM International; 2011.
 - [18] ASTM D 613-15. Standard Test Method for Cetane Number of Diesel Fuel Oil. West Conshohocken (PA): ASTM International; 2015.
 - [19] Knothe G. Analyzing biodiesel: standards and other methods. *J Am Oil Chem Soc* 2006;83:823–33.
 - [20] Santos Jr. VO, Oliveira FCC, Lima DG, Petry AC, Garcia E, Suarez PAZ, et al. A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis. *Anal Chim Acta* 2005;547:188–96.
 - [21] Gontijo LC, Guimarães E, Mitsutake H, Santana FB, Santos DQ, Borges Neto W. Quantification of soybean biodiesels in diesel blends according to ASTM E1655 using mid-infrared spectroscopy and multivariate calibration. *Fuel* 2014;117:1111–4.
 - [22] Oliveira JS, Montalvão R, Daher L, Suarez PAZ, Rubim JC. Determination of methyl ester contents in biodiesel blends by FTIR-ATR and FTNIR spectroscopies. *Talanta* 2006;69:1278–84.
 - [23] Alves JCL, Poppi RJ. Quantification of conventional and advanced biofuels contents in diesel fuel blends using near-infrared spectroscopy and multivariate calibration. *Fuel* 2016;165:379–88.
 - [24] Souza LM, Mitsutake H, Gontijo LC, Borges Neto W. Quantification of residual automotive lubricant oil as an adulterant in Brazilian S-10 diesel using MIR spectroscopy and PLS. *Fuel* 2014;130:257–62.
 - [25] Alves JCL, Poppi RJ. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). *Talanta* 2013;104:155–61.
 - [26] Balabin RM, Lomakina EI, Safieva RZ. Neural network (ANN) approach to biodiesel analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. *Fuel* 2011;90:2007–15.
 - [27] Oliveira RR, Lima KMG, Tauler R, de Juan A. Application of correlation constrained multivariate curve resolution alternating least-squares methods for determination of compounds of interest in biodiesel blends using NIR and UV–visible spectroscopic data. *Talanta* 2014;125:233–41.
 - [28] Guimarães E, Mitsutake H, Gontijo LC, Santana FB, Santos DQ, Borges Neto W. Infrared spectroscopy and multivariate calibration for quantification of soybean oil as adulterant in biodiesel fuels. *J Am Oil Chem Soc* 2015;92:777–82.
 - [29] Mitsutake H, Guimarães E, Freitas HC, Gontijo LC, Santos DQ, Borges Neto W. Multivariate control charts based on NAS and mid-infrared spectroscopy for quality control of B5 blends of methyl soybean biodiesel in diesel. *J Chemom* 2015;29:411–9.
 - [30] Mazivila SJ, Gontijo LC, Santana FB, Mitsutake H, Santos DQ, Borges Neto W. Fast detection of adulterants/contaminants in biodiesel/diesel blend (B5) employing mid-infrared spectroscopy and PLS-DA. *Energy Fuels* 2015;29:227–32.
 - [31] Mazivila SJ, Souza LM, Abubakar IM, Mavulula EL, Borges Neto W. Determination of residual automotive lubricant oil and residual solvent used in a dry wash as adulterants in Brazilian S-10 diesel (B7) using mid-infrared spectroscopy data and chemometric methods. *Anal Methods* 2016;8:5427–34.
 - [32] Siteo BV, Máquina ADV, Santana FB, Gontijo LC, Santos DQ, Borges Neto W. Monitoring of biodiesel content and adulterant presence in methyl and ethyl biodiesels of jatropha in blends with mineral diesel using MIR spectrometry and multivariate control charts. *Fuel* 2017;191:290–9.
 - [33] Mogollon NGS, Ribeiro FAL, Lopez MM, Hantao LW, Poppi RJ, Augusto F. Quantitative analysis of biodiesel in blends of biodiesel and conventional diesel by comprehensive two-dimensional gas chromatography and multivariate curve resolution. *Anal Chim Acta* 2013;796:130–6.
 - [34] del Río V, Larrechi MS, Callao MP. Sequential injection titration method using second-order signals: determination of acidity in plant oils and biodiesel samples. *Talanta* 2010;81:1572–7.
 - [35] de Juan A, Tauler R. Chemometrics applied to unravel multicomponent processes and mixtures: revisiting latest trends in multivariate resolution. *Anal Chim Acta* 2003;500:195–210.
 - [36] de Juan A, Tauler R. Multivariate curve resolution (MCR) from 2000: progress in concepts and applications. *Crit Rev Anal Chem* 2006;36:163–76.
 - [37] de Juan A, Tauler R. Factor analysis of hyphenated chromatographic data: exploration, resolution and quantification of multicomponent systems. *J Chromatogr A* 2007;1158:184–95.
 - [38] Neves ACO, Tauler R, Lima KMG. Area correlation constraint for the MCR-ALS quantification of cholesterol using EEM fluorescence data: a new approach. *Anal Chim Acta* 2016;937:21–8.
 - [39] Dantas WFC, Alves JCL, Poppi RJ. MCR-ALS with correlation constraint and Raman spectroscopy for identification and quantification of biofuels and adulterants in petroleum diesel. *Chemometr Intell Lab Syst* 2017. In Press.
 - [40] ASTM E203-16. Standard Test Method for Water Using Volumetric Karl Fischer Titration. West Conshohocken (PA): ASTM International; 2016.
 - [41] Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: new features and applications. *Chemometr Intell Lab Syst* 2015;140:1–12.
 - [42] Kennard RW, Stone LA. Computer aided design of experiments. *Technometrics* 1969;11:137–48.
 - [43] Tauler R, Smilde A, Kowalski B. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J Chemom* 1995;9:31–58.
 - [44] Tauler R, Kowalski B, Fleming S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal Chem* 1993;65:2040–7.
 - [45] Jaumot J, Gargallo R, de Juan A, Tauler R. A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB. *Chemometr Intell Lab Syst* 2005;76:101–10.
 - [46] Jaumot J, Igne B, Anderso CA, Drennen JK, de Juan A. Blending process modeling and control by multivariate curve resolution. *Talanta* 2013;117:492–504.
 - [47] Bro R, de Jong S. A fast non-negativity-constrained least squares algorithm. *J Chemom* 1997;11:393–401.
 - [48] Brereton RG. *Chemometrics for pattern recognition*. first ed. Chichester: John Wiley & Sons Ltd; 2009.
 - [49] Geladi P, Kowalski BR. Partial least squares regression: a tutorial. *Anal Chim Acta* 1986;185:1–17.
 - [50] Naes T, Isaksson T, Fearn T, Davies T. A user-friendly guide to multivariate calibration and classification. first ed. Charlton: NIR publications; 2002.
 - [51] Olivieri AC, Faber NM, Ferré J, Boqué R, Kalivas JH, Mark H. Uncertainty estimation and figures of merit for multivariate calibration. *Pure Appl Chem* 2006;78:633–61.
 - [52] Oliveira FM, Carvalho LS, Teixeira LSG, Fontes CH, Lima KMG, Câmara ABF, et al. Predicting cetane index, flash point and content sulfur of diesel-biodiesel blend using an artificial neural network model. *Energy Fuels* 2017;31(4):3913–20.
 - [53] Gambarra Neto FF, Marino G, Araújo MCU, Galvão RKH, Pontes MJC, de Medeiros EP, et al. Classification of edible vegetable oils using square wave voltammetry with multivariate data analysis. *Talanta* 2009;77:1660–6.
 - [54] Pasquini C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *J Braz Chem Soc* 2003;14:198–219.
 - [55] Workman Jr. J, Weyer L. *Practical guide to interpretive near-infrared spectroscopy*. first ed. Boca Raton: CRC Press; 2008.
 - [56] Silverstein RM, Webster FX, Kiemle DJ. *Identificação espectrométrica de compostos orgânicos*. seventh ed. Rio de Janeiro: LTC; 2007.
 - [57] Yang C, Yang Z, Zhang G, Hollebone B, Landriault M, Wang Z, et al. Characterization and differentiation of chemical fingerprints of virgin and used lubricating oils for identification of contamination or adulteration sources. *Fuel* 2016;163:271–81.